УДК 004.6 doi: 10.26906/SUNZ.2025.1.124-128

Artem Protsenko, Volodymyr Fedorchenko

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

MOBILE APPLICATION SECURITY ANALYSIS MODEL BASED ON ARTIFICIAL INTELLIGENCE

Abstract. The article considers the possibilities of using neural networks to ensure a secure environment for using devices. It reviews several neural network architectures that are already used to prevent attacks by attackers, the main areas of attack on mobile applications, and learning algorithms. It describes the features of using recurrent neural networks to analyze the dangerous space. The results of the article show that neural networks can be an effective tool for preventing data loss and hacker attacks. However, further research is needed to optimize the architecture and parameters of neural networks to improve the accuracy of threat detection.

Keywords: neural network, ML, DL, ANN, mobile application, cyber threat, firewall.

.

Introduction

Statement of the problem. The use of mobile devices has recently become an integral part of our lives. Today, they contain a lot of private information, such as banking data or documents. Many applications do not have proper data protection systems that can give the user confidence in the confidentiality of data. One of the possible solutions may be neural networks. The development of the field of artificial intelligence can help in the development of security methods for any programs used everywhere.

Analysis of recent research and publications. In the work of Draguntsov R., Rabchun D., Brzhevska Z. [1], the main vectors of attacks on information systems where Android OS applications are used as client interfaces are considered, compared and analyzed. This analysis is carried out in order to obtain initial material for the creation of practical methods for ensuring security at the architecture level of such systems. The article divides into categories possible attacks and vulnerabilities that underlie them, in the context of Android application security and taking into account the security model of the operating system itself and the environment.

In the work of Sosnovy V.O. and Zamriy I.V. [3], the effectiveness of recurrent neural networks for protecting information systems in cyberspace was investigated. The experiment showed that RNN works much better than basic machine learning algorithms. This is possible due to the built-in memory of RNN, where it is possible to remember several previous states and implicitly separate characteristic features, hidden complex structure and complex sequential communication within the data, which helps to achieve the best accuracy.

In the work of Ivanichenko Yu., Sablina M., Kravchuk [4], the main machine learning technologies that can be implemented in the cybersecurity organization are formulated. The main type of artificial neural network that will be used to prevent and detect cyber threats is also described, and it is established that the main machine learning technology for general application is artificial neural networks based on a multilayer perceptron with backpropagation of errors.

In the work of Lakhno V., Erbolat K., Bagdat Yu., Kryvoruchko O., Desyatko A., Tsyutsyura S., Tsyutsyura M. [7] a new approach to improving the information security of the educational institution network is proposed.

The proposed approach is structured and systematic. It allows to assess the degree of security of the educational institution network as a whole, as well as its individual subsystems and individual components that ensure the information security of the educational institution. Heuristic, expert, statistical and other indicators are used to assess the security of the system. The proposed model will allow to describe the process of ensuring the information security of the university network. The authors proposed a balanced system of IS indicators, which allows to assess the effectiveness of methods for ensuring the security of the university network in real time.

Purpose of the article. Creating a model for analyzing the security of a mobile application based on artificial intelligence technology. Review of methods for improving vulnerability recognition using neural network tools. Review of the use of neural networks of various architectures to prevent data leaks through unverified services, and the destruction of software design due to viruses entering the system.

Presentation of the main material

Every app store user data. It is this data that users try to hide from attackers when using their devices. This data should be properly protected so that other apps cannot access it, but this protection is not always provided.

JNI. Java Native Interface (JNI) is a mechanism in Java that allows you to interact with code written in other programming languages, such as C or C++. This is especially useful in cases where you need to call functions written in a low-level language or use libraries written in C languages.

Security flaws in the Java Native Interface (JNI) in Android applications can have devastating consequences, leading to vulnerabilities, attacks, or unauthorized access to some critical device resources. Here are some types of issues and how to fix them:

- Security vulnerabilities in the code C/C++:

Impact: The threat of memory leaks, buffer overflows, and unsafe function calls may lead to uncontrolled application behavior, including the possibility of malicious code execution.

- Incorrect handling of pointer data:

Impact: Faulty pointers or incorrect memory operation may lead to confidential data leakage or device damage.

- Insufficient permission checking:

Impact: Using insufficient permission checks may lead to unauthorized access to functions or resources, resulting in security vulnerabilities.

- Leakage of confidential data:

Impact: Incorrect handling of sensitive data during transfer between Java and C/C++ code may lead to data leakage, with serious implications for user privacy.

- Insecure call from Java to C/C++ code:

Impact: Lack of proper control and validation when calling functions from Java may lead to unpredictable behavior or vulnerability attacks.

- Using untrusted libraries:

Consequence: Using libraries with security vulnerabilities can compromise the entire application.

To minimize risks and increase security when using JNI, it is important to follow security programming best practices, carefully review C/C++ code, use secure functions, and implement access control and permission checks.

Danger areas of mobile applications. When analyzing mobile applications, first of all, it is necessary to outline the main areas of attack that can be carried out on the following systems:

- on server-side processes.
- on the client application.
- on the communication channel.

Server-side attack vectors. The information system, which involves the use of a mobile application as a client interface, contains a server part that implements the main part of the functionality free of charge.

The processes of the server part of the program are understood as any processes in the information system that are provided by functional, implemented server software. Such elements include:

- authentication, authorization and access delimitation.
- session control.
- business logic.
- data validation.
- error handling.

Attack vectors on the client application. Attacks carried out on a client mobile application within the studied architecture can primarily affect data that is processed locally. The main functional elements specific to the client application include:

- user interface.
- interprocess communication.
- local data storage.

Attack vectors on the communication channel. Many modern client-server applications for Android OS require HTTP communication protocol. Using this protocol without additional protection can be considered a serious vulnerability of the system, as there is no way to ensure the confidentiality and integrity of information during transmission.

To provide these features, a secure connection protocol - HTTPS is used. However, given the fact that mobile devices mostly operate in an untrusted environment that can be controlled by a potential attacker, the possibilities of using the regular HTTPS protocol are limited.

From this, four possible types of attacks on the communication channel can be distinguished:

- MitM attack on an unsecured communication channel (HTTP).
- MitM attack with DNS server substitution on the network on a secure communication channel (HTTPS).
- exploitation of vulnerabilities in the implementation of SSLPinning technology to attack a secure channel.
- cryptographic attacks on a secure communication channel.

Neural networks. Machine learning (ML) is the process by which machines learn from information they are given, building logic and predicting a certain output for a given input. There are three types of ML: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning uses a dataset labeled with a correct answer symbol to learn from. These labels define characteristics for each dataset. Once the model has completed its self-learning, it can begin to make predictions or decisions about new data or situations that are subsequently fed to it.

In unsupervised learning, there is no need for such a defined data set. When a model is given a data set, it automatically finds patterns and relationships, creating clusters within them. Unfortunately, this type of learning cannot predict anything. When new data is added, the model assigns it to one of the existing clusters or creates a new one. Reinforcement learning is the ability of a system to interact with its environment and determine the best results.

The system is “rewarded” or “punished” with a score for correct or incorrect answers, and based on the positive reward results obtained, a model is automatically formed. Similarly, after training, an artificial neural network is prepared to predict new data that is given to it. Deep learning (DL) is a class of ML algorithms that uses multiple layers to gradually extract higher-level features from the original input.

The main differences between ML and DL are ML algorithms almost always require structured data, while DL networks rely on an artificial neural network (ANN) layer. Quite often, ML requires human intervention to produce further results with a larger data set, while DL does not. One of the core concepts of DL is ANN.

ANN is a model that is based on the principle of organization and functioning of the human brain (i.e., a network of nerve cells in a living organism). In other words, a neural network algorithm attempts to create a function that maps input data to desired results. Neural networks (NNs) are typically organized into layers (Figure 1). Layers consist of a set of interconnected "nodes" that contain an "activation function." Patterns enter the network through an "input layer," which passes data to one or more "hidden layers," where the actual processing is done through a system of weighted "connections." The hidden layers are then fed to an "output layer," where the response is the output.

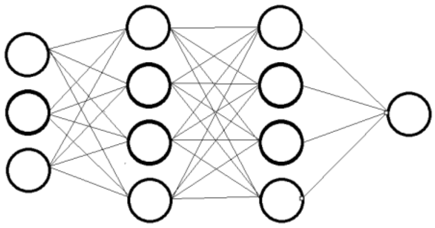

Fig. 1. Example of a neural network

Neural networks in cybersecurity. The scope of ML in cybersecurity is vast: from detecting anomalies and suspicious or unusual behavior to identifying your own vulnerabilities and fixing known ones.

For example, Reathi and Malathi published a set of ML algorithms trained on the NSL-KDD intrusion detection dataset for detecting abuse. Meanwhile, Buchak et al. focused on network intrusion detection using ML [6].

Meliher et al. proposed using NN for password strength verification. They compressed their model to hundreds of kilobytes and developed a JavaScript-like tool on the client side. For the active phase of password strength verification, they used neural networks such as Multilayer Perceptron (MLP) and Single Layer Perceptron (SLP). It should be noted that MLP provides better results than SLP when testing datasets. In addition, the number of layers is 10, and thus a better result is obtained. User and Subject Behavior Analysis (UEBA) uses the capabilities of ML to analyze behavioral patterns and network traffic in real time and respond instantly to attacks. This process can be accomplished by forcing the user to log in again, blocking the attack, or assessing the risk level and alerting the company's information security officers so that they can take the necessary measures.

Most ML and DL methods, such as clustering and decision tree learning, are used to detect abuse, anomalies, and hybrid cyber intrusions.

DARPA is working with BAE Systems to develop a system that will allow sensors to be configured and defenses to be deployed "at machine speed." The initiative, called CHASE, stands for Cyber Hunting at Scale. It aims to create automated tools to detect and profile new attack vectors, gather the right contextual data, and distribute defenses both within and across companies [8].

The cyberattacks considered by hackers are related to the general idea of big news. Information that can be gathered from a media source can help predict similar incidents using NLP and ML methods.

We can also use ML to identify developers' programs. Rachel Greenstadt and Eileen Kaliskan have created a system that can "de-anonymize" programmers by analyzing source code or compiled binaries.

Another method of monitoring a system and network for malicious activity or policy violations is an intrusion detection system (IDS). Another method is an intrusion prevention system (IPS) - this system works in conjunction with an IDS. These systems detect intrusions and stop the detected fraudulent activity.

Both systems use supervised and unsupervised ML techniques to detect anomalies: point, contextual, and collective. The main task of a firewall is to ensure the network security of a system by controlling incoming and outgoing network traffic. Firewalls allow or block traffic by comparing its characteristics with predefined patterns (i.e., firewall rules).

In their paper, Ukar and Ozhan presented the results of automatic anomaly detection in firewall pattern repositories based on ML and high-performance computing methods such as Naive Bayes, kNN, table solutions, and HyperPipes. Firewalls filter content between servers and are also a method specifically designed for web application content. A web application firewall (WAF) is deployed in front of web applications. Thus, it analyzes bidirectional web interface (HTTP) traffic, and then shows and blocks any malicious or illegal content. To implement such functionality in WAF, developers choose regular expressions, tokens, behavioral analysis, reputation analysis, and ML technologies.

Among ML techniques, special predictive techniques can be used for data loss prevention (DLP) to minimize the risk of hacking or leakage. Practical DLP solutions allow us to set special rules that classify confidential and non-confidential information so that it cannot be disclosed fraudulently or accidentally by unauthorized users. This process can be created using supervised learning algorithms and two types of examples: positive examples (i.e. data that needs to be protected) and counterexamples (i.e. documents like the positive set but do not need protection).

Using ML in cyberattacks. This section describes how the process of finding cyberattack results can be achieved using ML. Automated vulnerability scanning is one of the most obvious and challenging tasks in cybersecurity. For example, CSRF attack files contain only 5% of the additives, as reported in the OWASP Top 10 2017, while most frameworks include CSRF protection.

Accordingly, Calzavaraetal introduced Mitch, the first ML-based tool for detecting the same CSRF black box, which identifies 35 new CSRF vulnerabilities in 20 websites out of 10,000 Alexa websites and three more tools for previously undetected CSRF vulnerabilities in working software that has already been tested. analyzed using another state-of-the-art tool. Mitch is a binary classifier that labels sensitive or insensitive queries using a random forest algorithm in a 49-dimensional feature space. Compared to the heuristic classifiers BEAP and CsFire, Mitch demonstrated the best F1 scores and accuracy (Table 1). Trustwave released an open-source scanning tool that uses facial recognition to automatically track topics on social media. Image scanning simplifies this process, removing false positives in search results and speeding up data analysis for the human operator.

Using the data obtained about the target, the attacker can attach pre-designed fake news to the victim.

Table 1 – **Validity measures of tested classifiers (BEAP, CsFire, Mitch)**

Classifier	Accuracy	Response time	F1
BEAP	0.29	0.88	0.44
CsFire	0.24	0.95	0.32
Mitch	0.79	0.65	0.71

ML tools help recognize fake news, but the researchers say that the best way for ML to do this is to learn to create such fake news on its own. So, they developed a supervised text generation model called Grover. The testing process used four classes of articles: human news, machine news, human propaganda, and machine propaganda. Amazon Mechanical Turk experts rated the skin condition, including their overall reliability. In the case of propaganda, the score increased from 2.19 for the manually created article to 2.42 for the machine-generated article.

SNAP_R was introduced at DEFCON 24. SNAP_R is the world’s first automated depth generator for deep-sea fishing on Twitter [6]. While previous tools are based on Markov chain models, SNAP_R is based on a finite NN with an LSTM architecture. Using Twitter as a medium offers some advantages for automated text generation. For example, it limits the length of a message, which reduces the likelihood of grammatical errors. Despite this, Twitter links are often shortened, making it very easy to mask malicious domains. This, in turn, has significantly increased the success rate of detecting such attacks from 5–14% on Markov chain-based tools to 30–66%, which supports 45% for manual fishing. In this case, attackers do not know the exact algorithm for detecting the malware but can understand certain features it uses by using developed test algorithms in the black box algorithm. MalGAN selection is a competitive network generative algorithm that generates examples of rogue malware that can mitigate black-box ML modeling. This element can minimize the level to almost zero, making it difficult for a retraining-based defense method to work against such examples. The architecture of MalGAN is shown in Fig. 2.

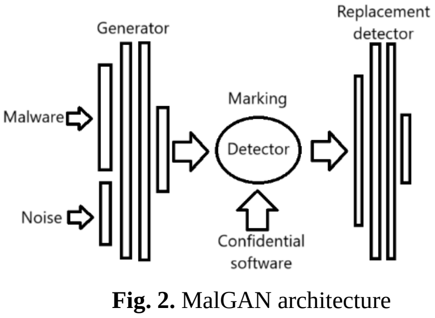

Fig. 2. MalGAN architecture

The generator uses the malware feature vector and the noise vector to transform the former into its generic version. The surrogate detector is needed to tune the black box detector and obtain gradient information to train the generator. Both networks are represented as multi-layer ANNs with direct input. The competing examples used against the black box detector come from a variety of machine learning methods trained on 160-dimensional binary vector functions representing API system calls, logistic regression, decision trees, support vector machines, and multilayer perceptrons, and voting on the ensemble of these algorithms. Malware developers retrain detectors after learning about such undetected examples, but MalGAN requires that this entire retraining epoch reach a 0% true positive rate.

Another example of the application of GANs in cybersecurity is password attack. There is a new method for generating passwords based on DL and generative adversarial networks, known as PassGAN. The key difference of this method is that DL does not require deep knowledge of the password structure, unlike methods based on templates, Markov models, and FLA. PassGAN helps improve the training of GAN Wasserstein (IWGAN) Gularajani et al. using the ADAM optimizer. The generator and discriminator in PassGAN are based on ResNets. The architecture of the generator and discriminator is shown in Fig. 3 and 4, and the residual block image is shown in Fig. 5.

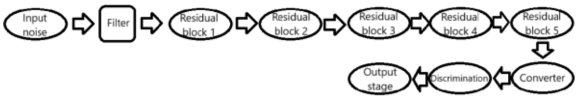

Fig. 3. Architecture of the PassGAN generator

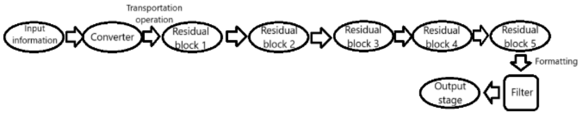

Fig. 4. PassGAN discriminative architecture

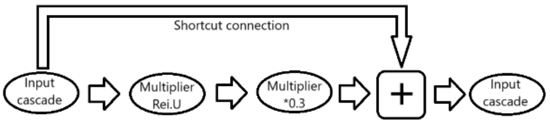

Fig. 5. Architecture of the residual block in PassGAN

To achieve maximum efficiency, attackers often choose multiple password cracking tools, such as HashCat, John the Ripper, PCFG, OMEN, and FLA, to combine various attack methods. For example, by combining raw PassGAN data with HashCat Best64 results, usability researchers found 51–73% more unique passwords than HashCat.

Traditional botnets wait for commands from a C&C, but now crooks are using automation to make decisions on their own. Fortinet developers predicted that cybercriminals would replace simple botnets with intelligent AI clusters of compromised devices, which have become a type of attack that can use self-learning to shape consumption of vulnerable systems with minimal oversight.

In the early stages of an attack, fraudsters often face the problem of missing a captcha. Suphannee et al. developed a simple attack that uses DL technology to semantically annotate an image. The system needs about 19 seconds per call to solve the tasks, with an accuracy of 70.78% for reCaptcha and 83.5% for Facebook image captcha. The system automatically selects which of the provided images are semantically like the sample images.

The system first obtains information for all images using Google ReverseImages Search (GRIS); Clarifai [6], which is built on deconvolutional networks; TDL, which is based on artificial Boltzmann machines or NeuralTalk and Caffe. Next, if no clue is found, the system searches the labeled dataset for images to obtain something like it, if possible.

Analysis model. Developing an artificial intelligence (AI)-based mobile application security analysis model is an important task due to the increasing number of mobile applications and the increasing threats to their security. Such a model can use various neural network techniques to identify potential threats and vulnerabilities in mobile applications. Here are some key aspects that can be considered when developing such a model:

- Source code analysis.
- Detection of abnormal behavior.
- Network activity monitoring.
- Malware detection.
- Vulnerability analysis in third-party libraries.
- Monitoring changes in system calls.
- Detecting social engineering attacks.
- Regular updates of rules and algorithms.

When choosing an AI training architecture, two options were considered:

- Tutored learning (perceptron).
- Unsupervised learning (adaptive resonance networks).

Studying with a teacher. Supervised learning is not suitable for this domain analysis because the perceptron has limited effectiveness when dealing with complex class boundaries and is not capable of solving nonlinear problems. This method is more suitable for simple projects. Blended learning is not suitable due to the high complexity of the training and very long computation time. This method is more suitable for very large projects where immense data sets need to be processed.

Studying without a teacher. Adaptive Resonance Theory (ART) networks are a type of neural network that can be used in cases where the data has a nonlinear structure or when it is necessary to consider the dynamic behavior of the data. ART networks can be used for clustering and timely detection of leaks and outliers.

Conclusions

The article examines the use of neural networks to ensure security using artificial intelligence. The latest research in this field is analyzed. It is demonstrated that neural networks can be an effective tool for solving security problems. The article examines the areas of danger of mobile applications, the use of neural networks in the field of cybersecurity. The results of existing AIs that cope well with the tasks assigned to them in detecting cyber threats are considered.

Based on the research conducted, the following conclusions can be drawn neural networks can be an effective tool for detecting and preventing threats in mobile applications; it is important to choose the right neural network architecture and configure its parameters.

In further research in this area, it is advisable to study the following issues: creating our own practical model for analyzing mobile application threats based on AI. The training will be carried out in Python using the TensorFlow library. The unsupervised learning method was chosen for security analysis because it meets the requirements for creating the AI we need.

References

1. Drahuntsov R., Rabchun D. and Brzhevska Z. (2020) “PRINCIPLES OF ENSURING SECURITY OF INFORMATION SYSTEM ARCHITECTURE BASED ON CLIENT APPLICATIONS FOR ANDROID OS”. Electronic professional scientific publication “Cybersecurity: education, science, technology”. No. 4 (8). P. 49-60.
2. Enck W., Ongtang M. and McDaniel P. (2009) “Understanding android security”. IEEE security & privacy. No. 1. P. 50-57.
3. Sosnovy V.O. and Zamriy I.V. (2022) “NETWORK SECURITY USING A RECURRENT NEURAL NETWORK”. Electronic specialist scientific publication “Word of a Scientist”. No. 5. P. 21-24.
4. Ivanichenko, Y., Sablina, M. and Kravchuk, K. (2021) “USING MACHINE LEARNING IN CYBER SECURITY”. Electronic professional scientific publication “Cybersecurity: education, science, technology”. No. 4 (12). P. 132-142.
5. Attack indicators based on artificial intelligence allow you to predict and stop threats as quickly as possible: website. URL: https://iitd.com.ua/news/
6. Sharma B., Mangrulkar R. (2019) “Deep learning applications in cyber security: a comprehensive review, challenges and prospects”. International Journal of Engineering Applied Sciences and Technology. No. 4(8). P. 148-159.
7. Lakhno V., Yerbolat K., Bagdat Y., Kryvoruchko O., Desiatko A., Tsiutsiura S. (2022). “Local network protection model of educational institution server virtualization system”. Cybersecurity: education, science, technology. No. 2 (18). P. 6-23.
8. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. Courville. (2017) “Improved training of Wasserstein GANs”. In Proc. of the 31st Int. Conf. on Neural Information Processing Systems. P. 5769-5779.

Received (Надійшла) 08.11.2024
Accepted for publication (Прийнята до друку) 12.02.2025

Модель аналізу безпеки мобільних застосунків на основі штучного інтелекту

А. С. Проценко, В. М. Федорченко

Анотація. Проводиться огляд декількох архітектур нейромереж, що вже використовуються для запобігання атак зловмисників, основних напрямів атак на мобільні застосунки, алгоритмів навчання. Описуються особливості використання рекурентних нейронних мереж для аналізу небезпечного простору. Результати статті показують, що нейронні мережі можуть бути ефективним інструментом запобігання втрат даних та хакерських атак. Однак необхідні подальші дослідження для оптимізації архітектури та параметрів нейронних мереж для покращення точності виявлення загроз.

Ключові слова: нейромережа, ML, DL, ANN, мобільний застосунок, кіберзагроза, брандмауер.