ЎЗБЕКИСТОН РЕСПУБЛИКАСИ АХБОРОТ ТЕХНОЛОГИЯЛАРИ ВА КОММУНИКАЦИЯЛАРИНИ РИВОЖЛАНТИРИШ ВАЗИРЛИГИ

МУҲАММАД АЛ-ХОРАЗМИЙ НОМИДАГИ ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ

ЎЗБЕКИСТОН РАДИОТЕХНИКА, ЭЛЕКТРОНИКА ВА АЛОҚА ИЛМИЙ-ТЕХНИКА ЖАМИЯТИ

ИҚТИСОДИЁТНИНГ РЕАЛ ТАРМОҚЛАРИНИ ИННОВАЦИОН РИВОЖЛАНИШИДА АХБОРОТ–КОММУНИКАЦИЯ ТЕХНОЛОГИЯЛАРИНИНГ АҲАМИЯТИ

Республика илмий-техник анжуманининг

МАЪРУЗАЛАР ТЎПЛАМИ
3-ҚИСМ

СБОРНИК ДОКЛАДОВ

Республиканской научно-технический конференции

ЗНАЧЕНИЕ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ В ИННОВАЦИОННОМ РАЗВИТИИ РЕАЛЬНЫХ ОТРАСЛЕЙ Э КОНОМИКИ

ЧАСТЬ 3

6-7 апрел 2017 йил

ТОШКЕНТ – 2017

DEEP PACKET INSPECTION TECHNOLOGIES

J.SH. Makhsudov (PhD, TUIT)
Sh.E. Narzullayev (TUIT, 1-course master)

The explosion of the commercial use of the Internet has created specific business and technology demands for products that could allow organizations to explore the opportunities that arose without compromising their security. Thousands of internal networks, with a high level of trust for their owners, have been connected to a public and loosely controlled network; this has opened those organizations to a series of new security problems.

One of the first concerns was the need of having a security mechanism that could allow basic definitions in terms of access control. The development of a network security policy to determine what resources could be accessed by which users, including the operations that could be performed, was always recommended as a good first step. Once the organization had this basic definition of the permissions that should be enforced at the connecting point with this new external world, it was ready to implement technologies for achieving this goal.

This vast use of HTTP and the other protocols that have been mentioned have forced most network and security administrators to create specific rules in their firewalls for allowing these types of communication in an almost unrestricted way. Several software developers of applications such as instant messaging or Internet telephony have adapted them for using these open communication channels, in an attempt to avoid organization enforced restrictions and controls. Some have even adapted their code to search and use any open port in the firewall, through approaches that remember port scanners, tools historically used for network and host security evaluation and invasion, although the reason for doing that can go beyond network security issues.

The network access control needs to become more granular, going beyond the basic functions provided by most technologies. The point is not blocking or not unblocking the HTTP port, but guaranteeing that this open port is being used only for specific types of authorized HTTP traffic. This includes protection against things like:

- Unauthorized download of mobile code, like ActiveX controls and Java applets
- Application-level attacks against Web sites
- Malware propagation through authorized protocols
- Use of authorized open ports by unauthorized applications
- Specific behaviors that could characterize an attack.

Different technologies have been used in these tasks, with limited success. Intrusion detection systems (IDS) were one of them. Although the main purpose of these technologies was to work as an auditing tool, several vendors have promised effective protection through firewall integration or active responses, such as connection resets. However, a Gartner report, published in 2003, pointed out several fundamental issues with the use of those systems, urging customers to replace them by new emerging technologies capable of not only detecting attacks, but blocking them in real time. Basically, the key arguments were:

- IDS cannot block attacks effectively, only detect them.
- Their detection capabilities were also limited, with a high number of false positives and negatives.
- The management burden is huge, theoretically demanding 24-hour monitoring of their functioning.
- They were not able to analyze traffic at transmission rates greater than 600 Mbps.

Although the report had some flaws, including technical errors like the speed limit, a huge and passionate debate was initiated. Security managers and professionals that have invested their budgets in IDS tried to justify their decisions. Vendors went even further, attempting to disqualify Gartner's arguments. But, curiously, most vendors at that time were already offering in their product ranges new options known as intrusion prevention systems (IPSs). These are

probably the most stable and mature technology capable of doing some of the actions demanded by the research report, which indicates that even they were aware of some of their product's limitations. Additionally, the report has also mentioned another recent Gartner research document that focused on a technology called deep packet inspection (DPI), that was new and then still loosely defined.

Deep packet inspection (DPI) is normally referred to as a technology that allows packet-inspecting devices, such as firewalls and IPS, to deeply analyze packet contents, including information from all seven layers of the OSI model. This analysis is also broader than common technologies because it combines techniques such as protocol anomaly detection and signature scanning, traditionally available in IDS and anti-virus solutions.It is right to affirm that DPI is a technology produced by the convergence of traditional approaches used in network security, but performed by different devices. The improvement of hardware platforms and the development of specific hardware devices for network security tasks have allowed functions that used to be carried out by separate components to be carried out by just one. However, it is not possible to argue that this convergence is complete. Vendors are still maturing their technologies and there is a huge space for improvement.Due to this convergence, it is important to understand which technologies have preceded DPI and what their drawbacks are because they have driven the demand for new technologies by not fulfilling all current network security needs.

One of the first technologies used for performing network security were packet-filtering firewalls. Those systems were implemented, basically, by using access control lists (ACL) embedded in routers. Access control was one of the primary concerns of the early age of commercial use of the Internet in the 1990s. Because routers are the connection point between internal and external networks, their use as access control devices were very natural and appropriate.

Simple packet filters analyze each of the packets passing through a firewall, matching a small part of their contents against previously defined groups of access control rules. In general, we can say that basic limitations were:

- Because they analyze individual packets, they could not identify security violations that can only be visualized by screening more of the traffic flow;
- Very little information from the packets was analyzed, avoiding the identification of several problems that could only be seen in the application layer.
- The rules were static, creating many security problems for screening protocols that negotiate part of the communication options, like ports and connections, on the fly (the FTP service is a classic example).
- In general, router ACLs, implemented through command-line parameters, are harder to manage than rules created in easy-to-use graphical user interfaces.

Due to those deficiencies, an alternative, known as application-layer firewalls or proxies, was developed. Designed with the purpose of solving the security limitations of the packet-filtering technology, proxies have adopted a very effective approach in terms of security, but are radical from the networking point of view.However, from a security perspective, pattern-matching approaches are even more ineffective in IDS than in anti-virus software. Most anti-virus software can block viruses in real-time once they are found, while most IDS can only generate an alert. They can also send a command to the firewall, asking for blocking of the source of a just-identified attack. However, this approach has at least two serious problems:

- Some attacks, including several denial-of-service techniques, can be performed using very few packets, disrupting their targets before the firewall responsible for blocking them receives any notification.
- IDSs are famous for their false positives. In case of a false alarm, the firewall can block legitimate traffic, compromising the availability of the services and creating huge administrative problems.

DPI technologies are based on a number of old approaches that used to be implemented by different devices. Hardware and software advances have allowed the convergence of those

approaches into single-box architectures that increases the security provided by them and makes their administration easier.However, single-box architectures lack defense in-depth, a key network security concept that has been used for years, which could lead to unnecessary exposure. Additionally, they create single points of failure that can compromise network availability. Nevertheless, both can be solved using technology largely available from most vendors and correct security design principles, implementing network perimeters according to specific security needs of each network. The popularization of the use of protocols with native encryption reduces the effectiveness of such solutions, but do not make then dispensable. Integrated approaches, using intrusion prevention controls, that normally include DPI, both at host and network levels, will probably be the best approach in the medium and long terms.

BASIC STRUCTURE OF ARTIFICIAL NEURAL NETWORKS

J.SH. Makhsudov (PhD, TUIT)
Sh.E. Narzullayev (master of TUIT)

Neural networks or connectionist systems are a computational approach used in computer science and other research disciplines, which is based on a large collection of neural units (artificial neurons), loosely mimicking the way a biological brain solves problems with large clusters of biological neurons connected by axons. Each neural unit is connected with many others, and links can be enforcing or inhibitory in their effect on the activation state of connected neural units. Each individual neural unit may have a summation function which combines the values of all its inputs together. There may be a threshold function or limiting function on each connection and on the unit itself, such that the signal must surpass the limit before propagating to other neurons. These systems are self-learning and trained, rather than explicitly programmed, and excel in areas where the solution or feature detection is difficult to express in a traditional computer program.

An Artificial Neural network (ANN) is a collection of neurons in a particular arrangement or configuration. It is a parallel processor that can compute or estimate any function. Basically in an ANN, knowledge is stored in memory as experience and is available for use at a future time. The weights associated with the output of each individual neuron represent the memory which stores the knowledge. These weights are also called inter-neuron connection strengths. Each neuron can function locally by itself, but when many neurons act together they can participate in approximating some function. The knowledge that is being stored will also be referred to as "numeric data", which is transferred between neurons through weights.

ANNs learn through training. There are various training algorithms for different types of ANN, based on the specific application. Based on the training, the weights associated with each neuron adjust themselves. By training, it is meant that a set of inputs is presented repeatedly to the ANN and the weights adjusted so that the weights reach an optimum value, where optimum means that weights either tend to minimize or maximize. An ANN is trained repeatedly and at one point it reaches a stage where it has 'learned' a particular desired function. With proper training it can generalize, so that even new data, which was not part of the training data, will yield the desired output.

When writing an ANN, this is mimicked by using a "perceptron" as the basic unit instead of the neuron . The perceptron can take several weighted inputs and summarize them, and if the combined input exceeds a threshold it will activate and send an output. Which output it sends is determined by the activation function and is often chosen to be between 0 and 1 or −1 and 1. Since the derivative of the activation function is often used in the training of the network, it is convenient if the derivative can be expressed in terms of the original function value, as few additional computations are needed to calculate the derivative in this case.

One problem with the ANN approach is over-fitting of the data, which happens when the classifier becomes to good at recognizing the training examples, at the expense of not being able

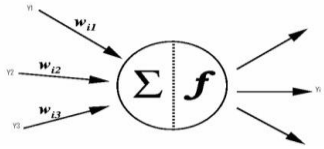
to recognize a general input. This can be avoided by cross-validation, where the network is trained on one set of data, and then evaluated on a separate one. When the error starts rising in the validation set, the network might be over-fitted. If previous networks are saved, the network can then be rolled back to the one which gave the smallest error.

The fundamental processing element of a neural network is a neuron. This building block of human awareness encompasses a few general capabilities. Basically, a biological neuron receives inputs from other sources, combines them in some way, performs a generally nonlinear operation on the result, and then outputs the final result.

Within humans there are many variations on this basic type of neuron, further complicating man's attempts at electrically replicating the process of thinking. Yet, all natural neurons have the same four basic components. These components are known by their biological names - dendrites, soma, axon, and synapses. Dendrites are hair-like extensions of the soma which act like input channels. These input channels receive their input through the synapses of other neurons. The soma then processes these incoming signals over time. The soma then turns that processed value into an output which is sent out to other neurons through the axon and the synapses.

Recent experimental data has provided further evidence that biological neurons are structurally more complex than the simplistic explanation above. They are significantly more complex than the existing artificial neurons that are built into today's artificial neural networks. As biology provides a better understanding of neurons, and as technology advances, network designers can continue to improve their systems by building upon man's understanding of the biological brain. But currently, the goal of artificial neural networks is not the grandiose recreation of the brain. On the contrary, neural network researchers are seeking an understanding of nature's capabilities for which people can engineer solutions to problems that have not been solved by traditional computing.

Neurons are basic building blocks of a neural network. They are processing elements that accept signal inputs, process them using a function and then form an output signal. They are also called basic nodes or units that accept inputs from an external source. Each input to a neuron has



a weight or gain associated with it. In Figure-1, the neuron output Y_i is a function of the weighted input that can be expressed by, $f\left(\sum_{j=1}^{a} W_{ij} * Y_j\right)$

W_{ij} are weights associated with the i W_{ij} th neuron where i=1, 2….n; j=1, 2, 3; n is the number of neurons. The Y_j are the inputs to the neuron. Y_i is the output of the i^{th} neuron. Figure-1 shows a simple neuron.

Figure-1. Simple neuron

A function is applied to the dot product of the weight vectors and input vectors. The dot product $W_{ij} * Y_j$ (for j = 1,2…n) is the net input to the i^{th} neuron. Also the output of the neuron could be an input to some other neuron through another set of weights.

These networks need to be capable of an infinite number of responses. Applications of this type include the "intelligence" behind robotic movements. This "intelligence" processes inputs and then creates outputs which actually cause some device to move. That movement can span an infinite number of very precise motions. These networks do indeed want to smooth their inputs which, due to limitations of sensors, comes in non-continuous bursts, say thirty times a second. To do that, they might accept these inputs, sum that data, and then produce an output by, for example, applying a hyperbolic tangent as a transfer function. In this manner, output values from the network are continuous and satisfy more real world interfaces. The greatest advantage of ANNs is their ability to be used as an arbitrary function approximation mechanism that learns

from observed data. However, using them is not so straightforward, and a relatively good understanding of the underlying theory is essential.

- Choice of model: This will depend on the data representation and the application. Overly complex models tend to lead to challenges in learning.
- Learning algorithm: There are numerous trade-offs between learning algorithms. Almost any algorithm will work well with the correct hyperparameters for training on a particular fixed data set. However, selecting and tuning an algorithm for training on unseen data require a significant amount of experimentation.
- Robustness: If the model, cost function and learning algorithm are selected appropriately, the resulting ANN can be extremely robust.

With the correct implementation, ANNs can be used naturally in online learning and large data set applications. Their simple implementation and the existence of mostly local dependencies exhibited in the structure allows for fast, parallel implementations in hardware.

THE POLITICAL POTENTIALS OF DPI

M.B. Mirzaeva (PhD, TUIT)

States have been monitoring and analyzing citizens' telecommunications since the telegraph, to the point of retaining encrypted text and banning certain modes of communications for fear that they would undermine state surveillance. DPI lets network operators monitor communications remotely and in real time for content of interest. Given its capacity to monitor the content of communications, DPI can be helpful in supporting 'lawful access' legislation and limiting the transmission of content the state has outlawed.

Lawful access legislation enhances policing and intelligence powers. There are typically three types of access powers associated with such legislation: search and seizure provisions, interception of private communications powers, and production of subscriber data. Deep packet inspection equipment is most useful in intercepting communications, and can be thought analogously as installing wiretap capabilities into digital networks. By installing DPI routers at key points in ISPs' networks, it is theoretically possible to remotely monitor communications of those suspected of engaging in illegal acts by making copies of all data traffic or specifically targeting one type of traffic (e.g. VoIP, web browsing, or peer-to-peer) and not logging or monitoring traffic that falls outside of the specified rule set. It is important to recognize that, while using DPI might be seen as the logical technology to facilitate state-based surveillance, this mode of monitoring differs from traditional wiretapping capabilities because of the breadth of communications that occur online. Whereas a traditional wiretap would capture voice communications, DPI-facilitated surveillance can capture and perform front-line analysis on any type of digital transaction, be it a voice communication, text-based chat, web browsing session, or any other kind of non-encrypted transmission. As such, DPI-based 'wiretapping' arguably stretches what it meant by wiretapping to a considerable degree, and it may not constitute 'maintenance' of state surveillance powers but an expansion of these powers.

As private copyright holders may be motivated to monitor for infringing files coursing across digital networks for civil reasons, the government may be concerned with monitoring and preventing content transmission it has deemed illegal. Using techniques similar to those exercised to monitor for copyright infringement, but with policies designed to take action on data traffic rather than just watching the wire for it, government could try to blacklist files known to contain child pornography, viruses, malware, disapproved encryption protocols, confidential or secret government documents, and so forth. Blocking or monitoring content could take the form of a government requiring certain routing equipment be installed in network providers' infrastructure or demanding that those same providers install and operate the equipment on the government's behalf. Moreover, such analysis and identification requires massively monitoring communications streams. Such actions do not focus on specific individuals, as with a wiretap.