УДК 004.738.5

05.13.19 - Методы и системы защиты информации, информационная безопасность (технические науки)

МЕТОДЫ ПОСТРОЕНИЯ ЭФФЕКТИВНЫХ АЛГОРИТМОВ ОБХОДА БЛОКИРОВОК ДОСТУПА К WEB-PECYPCAM НА ОСНОВЕ НТТР-ТУННЕЛЕЙ

Петров Антон Александрович к.т.н., доцент, SPIN-код: 2062-0236 ФГБОУ ВПО «Кубанский государственный аграрный университет», г. Краснодар, Россия

В настоящее время значительно набирает актуальность проблематика блокирования сетевого трафика в компьютерных сетях, в частности, в глобальной сети Интернет. Понимание механизмов блокировки, открытие новых возможностей по обходу такой защиты даст нам, с одной стороны, как новые методы и методики защиты от нежелательных блокировок, так и вектор развития средств, собственно, остановки нежелательного трафика, препятствия его прохождение через сетевые узлы. В статье предложен подход эффективного обхода ограничений доступа к вебсайтам на основе НТТР-туннелирования с минимальными затратами, совмещенный с подходом, применяющимся в пиринговых сетях. Дается описание алгоритма и его ключевых особенностей. Введен новый подход в задаче обеспечения доступности веб-сайтов, обладающий рядом преимуществ и устраняющий недостатки уже существующих решений в виде анонимных сетей - использование специальных узловых решений и сильная зависимость от количества узлов анонимной сети. Уровень развития современной сетевой инфраструктуры, переход общества и большей части человеческой цивилизации к информационно-сетевому существованию диктует нам новые требования к техническим средствам контроля и анализа сетевого трафик. Что, в свою очередь, требует разработки новых подходов к решению соответствующих задач

Ключевые слова: АНОНИМНОСТЬ, ДОСТУПНОСТЬ, ИНФОРМАЦИЯ, НТТР, ВЕБ-САЙТ, ПРОКСИ, ТУННЕЛИРОВАНИЕ

DOI: http://dx.doi.org/10.21515/1990-4665-153-025

UDC 004.738.5

Methods and systems of information protection, information security

METHODS OF CONSTRUCTING EFFICIENT ALGORITHMS BYPASSING BLOCKING ACCESS TO WEB-RESOURCES BASED ON HTTP TUNNELS

Petrov Anton Alexandrovich Cand.Tech.Sci., Associate Professor RSCI SPIN-code: 2062-0236 Kuban state agrarian University, Krasnodar, Russia

At present, the problems of blocking network traffic in computer networks, in particular, in the global Internet, are becoming increasingly topical. Understanding blocking mechanisms, opening new opportunities to circumvent such protection will give us, on the one hand, both new methods and methods of protection against unwanted blockages, and the vector of development of funds, in fact, stopping unwanted traffic, obstacles passing through network nodes. The article proposes an approach to effectively bypass access restrictions to websites based on HTTP tunneling with minimal costs, combined with the approach used in peer-to-peer networks. A description of the algorithm and its key features. A new approach has been introduced in the task of ensuring the accessibility of websites, which has several advantages and eliminates the disadvantages of existing solutions in the form of anonymous networks - the use of special node solutions and a strong dependence on the number of nodes in the anonymous network. The level of development of modern network infrastructure, the transition of society and most of human civilization to the information-network existence dictates new requirements for technical means of monitoring and analyzing network traffic. That, in turn, requires the development of new approaches to solving relevant problems

Keywords: ANONYMITY, ACCESSIBILITY, INFORMATION, HTTP, WEB SITE, PROXY, TUNNELING

1. Preamble

The new stage of which has come in the development of the exchange of information, characterized by intensive introduction of modern information technology, widespread local, corporate and global networks, creating new opportunities for information exchange.

Peer network (PS) are now one of the dominant technologies of the Internet. These networks provide a good decentralization - the lack of a single point of control and, most importantly, single points of failure. In other words - the service can be considered fully decentralized when to run it simply run the application without further input any data connection. Zero trust means that attacks such as MITM (Man-In-The-Middle) should, in principle, is not feasible due to the implementation of the Protocol (The network automatically reject malicious node, the network can function even if all nodes except you and to whom you pass captured, it is impossible to physically connect directly to the channel gap).

Many modern services in the Internet and Intranet operate on the basis of a peer network architecture. These are well-known P2P-services like Skype, BitTorrent, Viber, Tox, Storj, Ripple, BitCoin and many others.

However, the application of these technologies is impossible without attention to the issues of information security because of threats to information security, the circulated in these networks. Obviously, modern approaches to data transmission via the Internet do not fully meet the requirements or provide for the protection of information, or ease of use by the end user.

Thus, from a practical point of view for a long time there was a need in the software (SW) (preferably open source) that provides text, voice and video, as

well as the transmission and storage of data over the Internet (TCP / IP) between different kinds of devices - computers and mobile devices, etc. Of course, one cannot speak of ensuring a high degree of security of information processed in such software without building relevant algorithms and models, as well as their strict mathematical justification.

However, the development of techniques and systems to protect information from the point of view of their use in non-classical, dynamic, decentralized systems and networks are now virtually no research aimed at solving the problem of providing the necessary level of data protection throughout the period of operation of these information structures.

Another important point is that the existing decentralized algorithms use several gears "from each to each," which is not scalable and efficient, and most importantly - safe. As a result, not all of the known solutions can be effectively used and resiliency peer networks. Special solutions for the existing P2P systems have some drawbacks (for example, the presence of a single point of network failure, the lack of order, high bandwidth utilization, traffic office, the lack of resources to ensure the authenticity of duplicates).

It is no secret that every day many states strengthen control over their citizens on the Internet. We will not list these countries, since the purpose of this article - is not a political, but a purely technical. We only note that the countries where the freedom of information (including freedom of expression) on the Internet subject to significant restrictions is growing. Of course, the state apparatus justifies its actions concern for the safety of its citizens, but often broken when it is much more freedom than protected.

2. Introduction

We define the basic quality of Internet access, towards which we aspire: anonymous (the user does not want to call themselves, to show their presence), privacy (the denial of the transmitted data to any third parties), accessibility (the ability to access any public resources, regardless of the resolution of someone else). Several looking ahead, we note that it is solving the problem of access to information and will be devoted to this article. Of course, to ensure absolute performance described trinity - there is a global and complex in every scientific and technical problem. And to solve this problem is only necessary so that the solution does not depend on the right of the field, where it will have to operate.

How does going monitoring of user actions (in this case we are interested in a particular case - an attempt to access some resources, servers, sites)? Just note that it may be both offenders and law-abiding users. Widely known for several major ways:

1. Targeted blocking of various network resources (hosts, the individual services, the server entirely), according to the State, showing illegal activity on the Internet or other networks, such as the intranet provider. It is used as an active opposition towards the user and towards life - the whole server or the site is closed.

2. If the resource is not the geographical territory of a backbone providers directive imposed on locking, an entry in the table of access routers.

3. The worst-case scenario, which is used only in a single country - China is blocking VPN-connections. Blocking, as opposed to one or two cases, already made at several levels of the network model - the network (IPSec), transport (individual ports are blocked), session (L2TP), etc.

Of course, to counter the threats listed above do not necessarily have to be an intruder. What if the site was blocked only because of some innocuous remarks that seemed wrong to government officials? Perhaps just the error occurred? The list is long, but the main problem is clear - more and more often occurs blocking of sites on the Internet. Again, we shall not give concrete examples of blocking sites with the date, country and time - the curious reader can not (yet) find a lot of these cases on the Internet.

3. Problem

Thus, we formulate the task. Assume that in a network are the most aggressive security policy: Prohibited virtual tunnel, is an active opposition and blocking technologies such as Tor, I2P, peer-to-peer and other. Ultimately, everything is permitted to the user, is the use of multiple application layer protocols such as HTTP, FTP. Moreover, active monitoring is carried out and cut off all requests to the banned sites. Required to overcome opposition from the network equipment, and gain access to arbitrary resources, in particular site [1-5].

Immediately determine if the network is physically cut off from the Internet, or disconnected common application protocols, the problem has no solution a priori.

To solve this problem, we will use the protocol HTTP - text application layer protocol.

4. Proposed algorithm

At the moment, the key players are the versions of HTTP [6, 7]:

1. HTTP / 1.1. The current version of the protocol, adopted in June 1999. New in this version was the regime of "permanent link»: TCP-connection can remain open after sending the reply to the request, which allows you to send multiple requests in a single connection. The client is now required to send the information about the host name to which it refers, which made possible a simpler organization of virtual hosting.

2. HTTP / 2. Unlike previous versions, the HTTP / 2 is binary. Among the key features multiplexing requests for prioritization of requests header compression, uploading multiple items in parallel by a single TCP connection, support proactive push-notifications from the server.

Each HTTP-message consists of three parts, which are transmitted in this order:

1. The start line (Eng. Starting line) - determines the type of message;

2. Headings (Eng. Headers) - characterize the body of the message, the transmission parameters and other information;

3. The body of the message (Eng. Message Body) - directly with the message data. It is sure to be separated from the header by a blank line.

Headers and the message body may be missing, but the starting line is a must, as indicates the type of request / response.

The key idea is to use a HTTP-tunnel, combined with the approach of peer to peer networks. Performs environment and the platform on which the system will work unit is web-server (nginX, Apache, or whatever), to carry out normal job of hosting some of the sites. However, it is running in daemon (language is not important, it may be even PHP), providing the required functionality. In fact, it provided the use of a decentralized overlay network protocol running over HTTPS - the possibility of building a chain from an HTTP proxy server.

First of all, note that multiple proxy can be arranged only through HTTPS proxy - that is, proxy support method CONNECT. Refer to the HTTP / 1.1 protocol will discover a method CONNECT, entered to support HTTPS (secure connection to the web server). The algorithm works as follows:

a) the proxy sent the connection request to the resource (remote socket);

b) If you are permitted (authentication, ...) proxy tries to connect to the specified resource;

c) If all was successful, a positive response is sent. Then go through the channel data between the browser and the remote resource;

Sample dialogue (request and response ends with a blank line, followed by the raw-data):

CONNECT 222.111.111.121:443 HTTP/1.1 Connection: Keep-alive Host: 222.111.111.121:443 HTTP/1.1 200 Connection established

<RAWDATA>

Proxy in this case acts as a bridge. General view of the HTTP proxy shown in Fig. 1.



Fig. 1. The process of HTTP proxy

We give a number of important comments to the structure of the work of the protocol:

• The part of the user agent acts as a browser.

• Select a specific anonymous proxy site specially selected module installed, the benefit of modern browsers are actively using the extension mechanism.

• Each proxy node anonymous network uses adaptive algorithms data statistics of the leads, it updates the database of other nodes, and the like.

• There is an opportunity not only to ensure the availability of sites that are blocked by the ISP, but the function of anonymity, as well as the transfer of data between arbitrary users through a chain of anonymous in the proposed network of proxy sites.

• The work comes through the protocol HTTPS, with the use of SSL or TLS.

The generalized algorithm of operation is shown in Fig. 1. Suppose that there are three networks, which are roughly divided between a boundary routers (it happens to them to control access to sites). In other words, the three segments of the Internet, under the control of three different countries.



Fig. 2. The general scheme of the system. P1, P2, P3 - respectively, a plurality of anonymizing proxy servers that make up the foundation of the system. Ua1, Ua2 - A network clients; Ra3 - resource (website) A network; Ub1, Ub2 - clients network B; Rb1 - resource (website) network B; Rc1, Rc2 - international network of resources, Uc1 - resource (website)

network C.

Let users Ub1 and Ub2 of B-segment denied access to resources (sites) Rc1 and Rc2 of the segment S. The users Ua1 and Ua2 Segment A of the contrary, access to C-segment is allowed, as shown in Fig. 2. Then the generalized algorithm of the system is as follows:

1. You are trying to connect the protocol HTTP (S) to the WEB-server directly, either directly through a proxy (it all depends on the settings).

2. Selects a suitable proxy agent. The location of this node can be a segment of the user (with a non-zero probability of deterioration of communication quality, because it will take at least another one proxy node in the chain), or any other segment of the Internet.

3. There is a chain of alignment proxy CONNECT method through a series of protocol HTTPS.

4. Follow GET request through the proxy chain built by the encapsulation of its own meta application (over HTTPS) protocol.

5. returns data to the user agent (browser). Return Address WEB-server of course be replaced with a false.

Of course, p.p.3-5 require considerable detail their implementation, however, because of their design and the study is still at an early stage, leaving them to their next job description.

By and large, to block the work of such a system can only be fully disabling WEB, blocking HTTP protocol in general, seems to us unlikely that [8-13].

5. Reliability prognosis of the proposed algorithm, based on the system operation determinant characteristic dynamics analysis

At establishing the choice of method for reliability analysis it should be considered:

- systems reliability research problem class;

- adequacy and completeness of the mathematical model of the reliability characteristics, models performance, the model use initial requirements and limitations;

- the method usability for the computer simulation, for the data measurement during testing of the operation reliability characteristics statistical analysis in order to attain the calculation results accuracy characteristics;

- relevant to reliability characteristics formalization, employing information security methods, including mathematical and software research.

The sequence of the system reliability methods analysis is as following:

- system identification (purpose, scope, functions, structure, composition, backup, system maintenance, operation mode, external interaction,

qualifications of staff and quality of the software tools used in technology system which are planned at production organization, during system manufacturing);

- purpose of the system appointed definitions (range and required reliability characteristics values, system operation quality criteria, consequences of rejections, failures and boundary conditions criteria);

- baseline data determination (obtaining and pre-processing of source data about the reliability of components and counterparts, calculating the elements reliability characteristics, reliability distribution by elements of the system);

- analysis of the system:

- qualitative analysis (identification of types of faults, failures of mechanisms and their implications for the system, analysis of the scheme functioning, analysis of maintenance and repair formats, building reliability system structural scheme);

- quantitative analysis (mathematical model construction for the considered system reliability characteristics consistency in order to obtain quantitative reliability characteristics by calculation or simulation, analysis of failures and sensitivity importance, assessment the feasibility of improving the performance of subsystems based on backup strategies and maintenance strategy);

- evaluation of analysis results (comparison with the desired reliability characteristics and/or guidelines and tools to ensure the required reliability characteristics, which may include design review, identification of weaknesses, imbalances of modes, parts with high-risk of malfunctioning replacement, development of alternative ways to enhance reliability, implementation of trade-off analysis and designs options evaluation) [14].

Below we consider the program for reliability ensuring as the major form of scientific and technical reliability problem-solving results implementation. During operation a workable system enables a performance of n required functions, and the performance of each function is described by relevant characteristics as a function of time. During the general formulation of the system reliability prognosis problem it is required to predict the dynamics of changes for each of these characteristics. But in some cases, among the sequences of these n characteristics, it may be justified only one of its as principal, or determinant, which characterizes the system reliability foremostly. Based on this statement, hereafter, we consider the task of system reliability prognosis within the frame of determinant characteristics dynamics changes analysis.

For such integral characteristics is proposed to use the characteristic formation (accumulation) of failures in the examined system. Dynamics of such characteristics change will be considered in this section.

Mathematical model of this characteristic is a random process $A(\omega, t)$, $\omega \in \Omega$, $t \in T$, which implementation at a fixed $\omega = \omega_1$ will be marked as A(t), $t \in T$.

Such process, for example, can be the Gaussian random process with independent increments.

System reliability prediction problem and its solution will be considered in this section under certain conditions and will have a methodical nature with attached illustrations about realizations of the studied stochastic process $A(\omega, t)$.

First, we consider the relevant assumptions for the problem of the system reliability prognosis. It is believed that a reasonable functioning of the determinant characteristic $A(\omega, t)$ that describes the mode of system operation, establishes the value of this characteristic that leads to the failure (ultimate state). Thus, for the implementation of observation over the process $A(\omega, t)$, that is the function of time A(t), it should be defined statistical estimations of average rate of change and the variation coefficient for its speed rate; this generally makes it possible to calculate (predict) the reliability characteristics for the system, without using the mode of failure (fracture) of the system.

Thus, for the system reliability predicting it must be specified the following information about the dynamics of change at the functioning of the system determinant characteristic as of a random process $A(\omega, t)$:

- model of degradation process (monotonic or non-monotonic character of implementation);

- threshold meaning of limited value changes for the determinant characteristics A_{max} ;

- initial value of the determinant characteristics A_0 ;

- average rate of the determinant parameter change within operation conditions *a*,

- variation coefficient (mean square root deviation) for the determinant characteristic ν change rate.

Basing on the specific implementations of such characteristic's analysis, or basing on a general analysis of the physical processes of degradation caused by the change of the determinant characteristics (wear, corrosion, et cetera) it can be formulated the dominant degradation process and defined its type (monotonic *DM* non-monotonic *DN*). This serves as a reason for the decision to accept for the distribution operating time to failure (ultimate state) mathematical model from the corresponding species of distribution (DM - or *DN* - distribution).

The following are examples of these reliability characteristics for a random process $A(\omega,t)$, that describes the relevant processes of degradation in the examined systems, with a prospect to use the obtained characteristics for the problems of its prediction solution.

Example 1. Below we consider a stochastic process with independent Gaussian stationary increments $A(\omega, t)$ and monotone implementations, charts

of which are shown in Fig.3. In this process, the model $A(\omega, t)$ is described in the form

$$A(\omega, t) = A_0 + \eta_{\mathcal{M}}(\omega, t) \tag{1}$$

where $\eta(\omega t)$ - random process which has the following characteristics: $a = \mathbf{M} \left\{ \frac{d\eta_{M}(\omega, t)}{dt} \right\}$ - the average rate at the corresponding time interval t of the

system uptime; a > 0; $\frac{d\eta_{M}(\omega, t)}{dt} \ge 0$, $v = \sqrt{\mathbf{D}\left(\frac{d\eta_{M}(\omega, t)}{dt}\right)}_{a}$ - variation

coefficient.



Fig. 3. Realizations charts of a random process with independent Gaussian stationary increments (monotone distribution)

For this case, the distribution function of operating time to failure (ultimate state) is:

$$F_{DM}(t) = \Phi\left(\frac{\alpha t + A_0 - A_{\max}}{\nu \sqrt{\alpha t (A_{\max} - A_0)}}\right).$$
(2)

Subsequently, the reliability characteristics of the system are as following: - mean time between failures (average resource):

$$T_{cp} = \left(1 + \frac{v^2}{2}\right) \frac{\left(A_{\max} - A_0\right)}{\alpha} , \qquad (3)$$

- probability of flawless operation in the interval [0, t]:

$$R(t) = \Phi\left(\frac{A_{\max} - A_0 - \alpha t}{\nu \sqrt{\alpha t (A_{\max} - A_0)}}\right).$$
(4)

Example 2. A stochastic process with independent Gaussian stationary increments and nonmonotonic distribution, implementations graphics of which are shown in Fig. 4. is described in the form

$$A(\omega,t) = A_0 + \eta_H(\omega,t) ,$$

where $\eta_{H}(\omega, t)$ - the random process that has characteristics: $\alpha = \mathbf{M} \left[\frac{d\eta_{H}(\omega, t)}{dt} \right]$

- average rate,
$$\alpha > 0$$
; $\frac{d\eta_H(\omega, t)}{dt} \ge 0$; $\nu = \mathbf{V} \left[\frac{d\eta_H(\omega, t)}{dt} \right]$ - variation coefficient.



Fig. 4. Charts of realizations of a random process with independent Gaussian stationary increments (non-monotonic distribution)

In this case, the distribution function of operating time to failure (ultimate state) is:

$$F_{DM}(t) = \Phi\left(\frac{\alpha t + A_0 - A_{\max}}{\nu \sqrt{\alpha t (A_{\max} - A_0)}}\right) + \exp\left(\frac{2}{\nu^2}\right) \Phi\left(-\frac{\alpha t + A_{\max} - A_0}{\nu \sqrt{\alpha t (A_{\max} - A_0)}}\right).$$
 (5)

Expressions for the evaluation of object reliability indices are as following:

- mean time to failure (average lifetime):

$$T_{cp} = \frac{A_{\max} - A_0}{\alpha} , \qquad (6)$$

- probability of flawless operation in the range [0, t]:

$$R(t) = \Phi\left(\frac{A_{\max} - A_0 - \alpha t}{\nu \sqrt{\alpha t (A_{\max} - A_0)}}\right) - \exp\left(\frac{2}{\nu^2}\right) \Phi\left(-\frac{\alpha t + A_{\max} - A_0}{\nu \sqrt{\alpha t (A_{\max} - A_0)}}\right).$$
(7)

These examples for characterization of reliability which can be used in prognosis problems are characteristic for a wide class of engineering systems, that include hardware, mechanical and electrical systems. For such complexes the processing of degradation (aging, wear and tear) is accumulated integrally, along the growth of time to failure *t*. Therefore, the most reasonable model for such physical processes is a stochastic process with independent increments. With regard to the large number of random factors in the degradation process, a typical model for increments distribution law with independent increments must be the Gauss law.

Determining such parameters values as the average rate of the process α change and the variation coefficientv, as well as determining the initial value A_0 and limiting A_{max} for the studied systems operation can be made by calculation, using the known methods of mathematical statistics. It should be mentioned significant intricacy that could be met at such data statistical treatment problem solving.

5. Conclusions and future prospects of research

Thus, the proposed approach to the task of ensuring the availability of a number of advantages and eliminates the main drawback of the existing solutions in the form of anonymous networks - use of special nodal solutions and a strong dependence on the number of units anonymous network. For the successful operation of our solution configured and running fairly standard web server with a static public ip-address and support the implementation on Python. Client enough to use standard browser (Chrome, Firefox, etc.) with the establishment of a special extension that makes a request for anonymous proxy network nodes and selects the optimal node based on the preset algorithm. Perform testing of performance and further improvements developed algorithms and models is expected in subsequent studies.

References

1. Juan Pablo Timpanaro, Isabelle Chrisment, Olivier Festor. Monitoring the I2P network // INRIA Nancy-Grand Est. — Henri Poincar´e University, France, 2011. — p. 5-7.

2. Adrian Crenshaw. Darknets and hidden servers: Identifying the true IP/network identity of I2P service hosts // In the Proceedings of Black Hat 2011. — Washington, DC, 2011.

3. D.e.I. Abou-Tair, L. Pimenidis, J. Schomburg, B. Westermann. Usability Inspection of Anonymity Networks. — Technical University of Dresden. — 2009. — 76 p. — ISBN ISSN 1430-211X.

4. Walls Colin. Embedded software. — Newnes, 2005. — P. 344. — ISBN 0-7506-7954-9.

Paul Buder, Daniel Heyne, and Martin Peter Stenzel. Performance of Tor. — Germany: Technische Universitat Darmstadt. — 15 p.

5. Mike Cardwell. Transparent Access to Tor Hidden Services // grepular.com.

6. HTTP/2 official standard // https://tools.ietf.org/html/draft-ietf-httpbis-http2-17

7. RFC 2616 // https://tools.ietf.org/html/rfc2616

8. Breaking the Collusion Detection Mechanism of MorphMix by Parisa Tabriz and Nikita Borisov.

In the Proceedings of the Sixth Workshop on Privacy Enhancing Technologies (PET 2006), Cambridge, UK, June 2006, pages 368-384.

9. Ignoring the Great Firewall of China by Richard Clayton, Steven J. Murdoch, and Robert N. M. Watson.

10. In the Proceedings of the Sixth Workshop on Privacy Enhancing Technologies (PET 2006), Cambridge, UK, June 2006, pages 20-35.

11. Linking Anonymous Transactions: The Consistent View Attack (PDF) (Cached: PDF)

by Andreas Pashalidis and Bernd Meyer. In the Proceedings of the Sixth Workshop on Privacy Enhancing Technologies (PET 2006), Cambridge, UK, June 2006, pages 384-392.

12. Privacy Implications of Performance-Based Peer Selection by Onion Routers: A Real-World Case Study using I2P by Michael Herrmann and Christian Grothoff. In the Proceedings of the 11th Privacy Enhancing Technologies Symposium (PETS 2011), Waterloo, Canada, July 2011.

13. Non-Discretionary Access Control for Decentralized Computing Systems by Paul A. Karger. Laboratory for Computer Science, Massachusetts Institute of Technology S. M. amp; E. E. thesis MIT/LCS/TR-179, May 1977. Chapter 11, "Limitations of End-to-End Encryption," has some early discussion of traffic analysis issues.

14. Reliability basics of information systems / ed. by Alexander PETROV ; [aut.] Alexander PETROV, Vladimir Khoroshko, Leonid Scherbak, Anton Petrov, Marek Aleksander. — Kraków : AGH University of Science and Technology Press, 2016. — 246, [1] s.. — Bibliogr. przy rozdz.. — ISBN: 978-83-7464-859-2