Empirical Study of New Metrics for the Internet Route Hijack Risk Assessment

© Vitalii Y. Zubok and © Igor Kotsiuba

Pukhov Institute for Modelling in Energy Engineering National Academy of Sciences of Ukraine, Kyiv, Ukraine
ipme@ipme.kiev.ua

Abstract. Possibility of dynamic routes change between nodes which are not physically connected is a key feature of the Internet routing. With two key concepts - one-hop forwarding in routing process and possibility of address space aggregation for routing purposes, the Internet became global and can grow virtually unlimited. However, one of the most significant problems of the Internet connectivity is caused by the Border Gateway Protocol (BGP) weaknesses – lack of verification of input routing data. It leads to the so-called route leaks and route hijacks. None of proposed and partially implemented upgrades and add-ons which are referred to as MANRS can deliver reliable defense against those types of attacks. Route hijack detection services are mainly provided by third-party services such as BGPMon. They track worldwide routes by tracing and keep track of route announcements in BGP, and notify the network administrator of suspicious events related to their prefixes based on routing information. And the main problem is that monitoring alert is post-mortem reaction when the routing accident has already happened or is happening. That's why it is necessary to learn how to manage risks arising from cyber attacks on global routing. Assessing the risks of route interception requires quantitative measurement of the impact of an attack on the routing distortion, and therefore, the breach of information security. This offers a way of exploring the topology of connections between Internet nodes to further solve the risk management task with topology methods. In previous papers we used the knowledge of the features of the Internet topology to find the relationship between topology and global routing vulnerability. One of the most important steps was to build a formal model of global Internet routing with formal description for objects, relations and processes of the Internet routingsuch as the IP address, address space, network prefix and their encapsulation, route, best path, and routing itself. In this paper we offer new node metrics for representation of both components of information security risk - possible losses and likelihood of losses. The first metric, which we have, called 'significance', is tied it to importance of node in routes distribution, with impact of number and weight of announced prefixes. The second metric, called 'trust', reflects likelihood of hijacking a route on a particular node. Finally, we demonstrate some empirical results of how these metrics can model the effective network topology regarding relaxation risks of route hijack.

Keywords: The Internet, Global Routing, Route Hijack, Trust Metrics, Cyber security.

1 Introduction

The exterior gateway protocol BGP 4 has been developed to deliver this feature, along with policies and procedures of inter-domain routing. Developed for the network of hundreds of nodes which rely on information from each other, decades later the same-BGP-4 is applied to tens of thousands of nodes and is crucially lacking routing data integrity. Nowadays there are over 80000 nodes called Autonomous Systems (AS) interconnected in some way and thus building the telecommunication network – the Internet [1]. Such large number of transit nodes and even larger number of links moves us from the theory of graphs to the theory of complex networks, where the study of the general properties of topology is preferred to the study of specific connections between nodes [2], [3]. This is the starting point of route forges, route hijacks, and other frauds with global impact [4].

Proactive counteraction mechanisms are suggested, such as Resource Public Key Infrastructure (RPKI) [5]. It's a part of the Internet Routing Registry system. This service provides a collective method to allow one network to filter another network's routes. The method's operation begins with cryptographic signing of the route origin. A Route Origin Authorization (ROA) is a cryptographically signed object that states which AS is authorized to originate a certain prefix. A ROA contains three informational elements: the AS Number that is authorized, the prefix that may be originated from the AS, and the maximum length of the prefix. However, such techniques are fully effective only in global deployment, and operators are reluctant to deploy them because of the associated technical and financial costs. For example, Telia, one of the Tier-I Internet backbone operators, announced that it was using RPKI for security in its internet routing infrastructure only since September, 2019.

2 Approach to the Problem

In the face of the impossibility of reliable protection against damage associated with an attack, it is necessary to learn how to manage risks arising from cyber attacks on global routing. For this purpose, we must use well-studied topological peculiarities of the Internet to find methods of routing attacks mitigation by a direct improvement of the connections between Internet nodes.

3 Analysis of Recent Research and Publications

Anti-hijack protection consists of two steps: detection and mitigation. RPKI mechanism with route origin validation is not sufficient to mitigate AS hijacking. Analysis of the mechanisms of the attack, depending on its objectives and options for its implementation is described in detail in [6]. Detection is mainly provided by third-party services such as BGPMon. They notify the network administrator of suspicious events related to their prefixes based on routing information. They track worldwide routes by tracing and keep track of route announcements in BGP. In the event of an incident, the affected networks begin to mitigate the consequences of the event, for example by announcing more specific prefixes to their networks or by requesting other AS to filter out false announcements. There are some other studies which offer mechanisms for route attack detection such as ARTEMIS [7] and PeerLock [8].

However, due to the combination of technological and practical deployment issues, existing reactive approaches are largely inadequate. In particular, the most advanced technologies have the following major problems:

- variety of types of routing attacks and combinations of methods leads to lack of a reliable method for detecting route interception;
- operators should be informed in advance of legitimate changes to their routing policy (new interactions between AS, announcement of a new prefix, etc.) so that such changes are not considered suspicious events for conditional third-party detection systems. Otherwise, adopting a less rigorous policy to compensate for the lack of updated information and reducing the number of false positives carries the risk of neglecting real events and not detecting false negatives;
- Only few minutes of unauthorized traffic diversion can result in heavy financial losses due to unavailability of service or security breaches. At the same time, the response time to incidents is slow in any case, as current practice requires the need to manually check alerts coming from monitoring systems and third-party services. Duration of widely known incidents ranged from several hours to months [4].

At the risk identification stage of risk assessment process, specific requirements to the quality of information are raised. There is a requirement of the highest possible level of completeness, accuracy and conformity at the time of its receipt. Quality requirements are also raised to the quality of information sources [9].

Our goal is as follows. Bearing the above-listed considerations in mind, we are trying to find the relationship between topology and routing vulnerability to obtain a method for quantifying information risk using a formal global routing model.

4 Basics of Global Internet Routing and the Nature of Route Hijack

Existence of links between Internet nodes is determined by existence of border interaction between groups of network communication equipment. With relation to border interaction, we suppose these groups as a node, or as an autonomous system. An Autonomous System (AS) is a group of IP networks having a single clearly defined routing policy which is run by one or more network operators. ASes exchange routing information with other ASes using Border Gateway Protocol BGP-4. Exterior routing decisions are frequently based on policy rules rather than purely on technical parameters [10]. A model of 4 AS interconnection is represented on a Fig. 1.

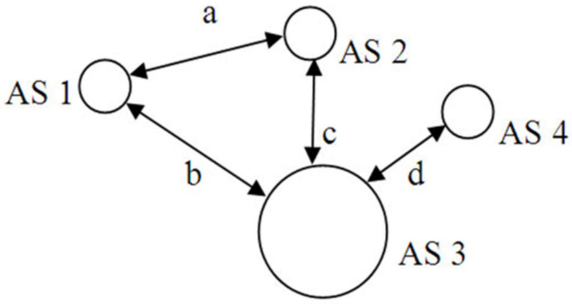

Fig. 1. Autonomous systems interconnection. AS 1,2,3,4 – autonomous systems; a, b, c, d – links

Each AS provides network prefixes to which it is ready to accept traffic, to a connected AS (it is called peer AS). So, AS4 has peering with AS3 using link d and announces its prefixes to AS3. It means that AS3 currently "knows" at least one way to transfer packets addressed to networks whose prefixes are announced by AS3. At the same time, AS3 announces to AS4 its prefixes too. As it's shown on fig. 1, AS3 also is peering with AS1 and AS2 using links b and c. Due to this, AS3 is able to re-announce AS4 prefixes accepted from AS1 and AS2, and vice versa, re-announce AS1's and AS2's prefixes to AS4. This ability comes from gateway protocol's features, and its presence is subject of a routing policy. Also, we can see that AS1 and AS2 are peering (a), so in AS1-AS2-AS3 triangle they are able to be a transit node for each other. However, AS4 has only one peer and it can't provide any transit. It's called "stub" AS.

Let us suppose, AS3 and AS4 are not linked, however AS3 by misconfiguration or maliciously announces to AS1 and (or) AS2 prefixes originated by AS4. Due to the lack of integrity inherent to BGP-4, AS1 and AS2 have no mechanisms to automatically verify and authorize those routes. More complex network of ASes is shown on Fig.2. AS6 is legitimate origin for 12.34.0.0/16 route, however if AS1 also announces this route, even in such easy network map we can see the nodes (AS2 and AS3) which accept this route as the best (shortest) path. Being aware that according to BGP model each BGP system can announce only one path –the best, i.e. shortest route for each prefix, we understand that AS2 and AS3 will use and propagate forged route to all their peers.

And let's look at Fig.3, where the route hijack has become a prefix hijack due to (erroneous or malicious) de-aggregation of 12.34.0.0/16 prefix to more specific 12.34.0.0/17 + 12.34.128.0/17; as a result, all other nodes will not use route to 12.34.0.0/16 because of existence of more specific ones. In this case affected ASes will not stop to announce legal route to whole prefix 12.34.0.0./16, although it can be used only if more specific /17 prefixes are not accepted by some AS for any reason.

When (or rather "if") the RPKI is implemented by 100% of Internet providers, including the largest Tier I networks, such hijack will not be possible due to route origin validation procedure, complementary to global routing. But there's nothing to counteract a man-in-the-middle attack with AS path forgery, when origin keeps looking valid (Fig.4).

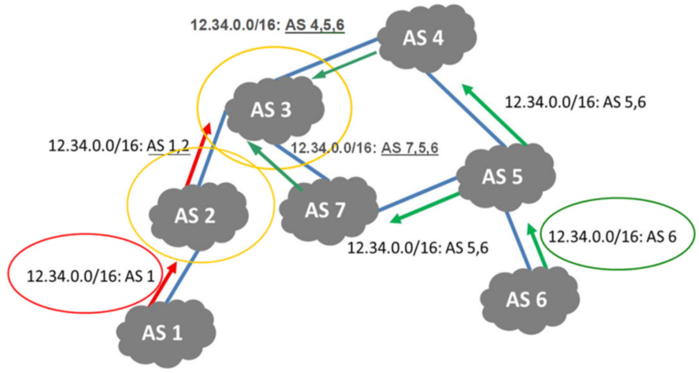

Fig. 2. AS1 performs hijack of the route to 12.34.0.0/16 belonging to AS6.

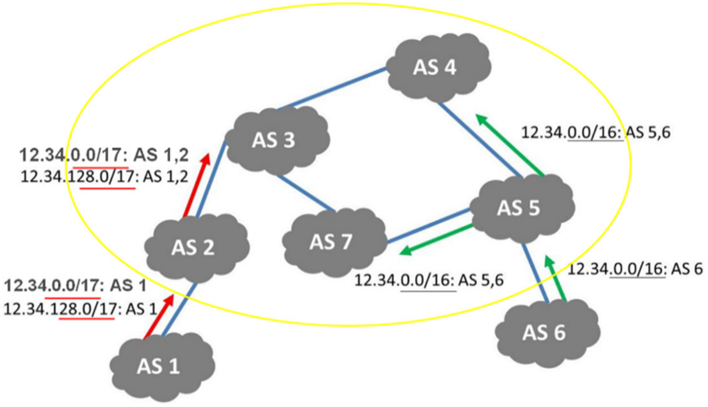

Fig. 3. AS1 uses deaggregating to hijack the route to 12.34.0.0/16 belonging to AS6.

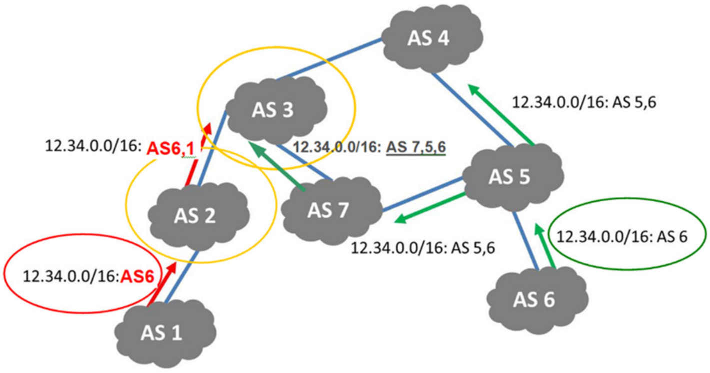

Fig. 4. AS1 forges route origin while AS2 is not using appropriate BGP filtering.

Any ideas of registering and validating complete set of legitimate Internet routes do not look realistic, both now and in future, for many reasons, including computing complexity and processing time. That's why we suppose, that global routing will be vulnerable for a long time.

5 Attempt of Formal Explanation of Route Hijack

Distance is the parameter routing attacks are tampering with. From a practical point of view, this means that a route is hijacked only if the distance through the fictitious route is be less than through the real route. So, let's find the formula for affecting the node with forged route. The task of finding the best route is complicated and non-linear. Therefore, the TCP/IP stack has adopted the so-called one-step approach to optimizing the packet route (next-hop routing) - each router and destination node only have to choose one step forward of packet transmission. A formal description of the Internet global routing objects and processes is described in [11]. Let us outline the process for choosing a prefix p(a) by destination IP address and then choosing a route with shortest path π (p):

$$p(a) = \{\min_j(p_j) \ : \ a \in p \subset A, \ 0 < j \le |A|\},$$

$$\pi_v(p) = \{\min_v(m_v(p)) : \pi \in M_p, v \in V_p\}.$$

For the sake of common case, we assume that our network is connected, that is, at least one route to any prefix is known at each node. If there are two or more of prefixes on a particular node u, BGP chooses one of them, based on known criteria, the most important of which is path length. After that, this route is in use at this node, and it will be announced to neighboring nodes. If at some node two or more routes have the same path length, the decision will be made according to secondary criteria. After passing each transit node, the route is extended by 1 node.

Consider at this stage the case of intercepting a route without de-aggregation. The hijack of prefix legitimately originated from node v, is that a spoofed route $\pi'(p_v)$ is announced to the network (typically from one particular node - [4]), competing with the true route $\pi'(p_v)$. In Figure 1, we can see that $\pi'(p_v)$ will obviously capture the nodes AS2 and AS3. On the other hand, AS4 and AS7 will receive a false route $\pi'(p_v)$ but it will lose to $\pi'(p_v)$. These nodes will not pass it on to their other neighbors.

In more complex topology we can see that on some hubs route hijack with initially one forged route can significantly increase the number of competing routes on some network hubs. At first sight, the most plausible way to model route distribution is methods of cellular automata, but this approach relies on periodic grid of cells, and we couldn't yet find the way to represent AS topology this way. However, it follows from the explanation of the hijack mechanism, that forged route leads to information risk only in two cases: (a) if it changes the route of IP packets to malicious node; (b) if it changes final destination of IP packets.

6 Introducing the New Node Metrics

As we described above, the inequality $\pi'(p_v) < \pi(p_v)$ for a particular node *v* is more likely under larger $d(v,u)$ - metric distance between nodes v and u. The extreme value of $d(v,u) = 1$ leads to impossibility to provide forged routes $\pi'(p_v)$ to node *v* if related true routes originated by node *u*. So, this should also eliminate for node *u* the risk of information security losses on node *v*.

For an intruder, it is easier to manipulate the path length if the path is longer. In a long path, in the middle there are more nodes through which one can announce a forged route. Therefore, the probability *P* of interception between nodes *u, v* increases for distant nodes and decreases for close ones:

$$P(v,u) \sim d(v,u).$$

Also, information losses increase with increasing number of affected nodes. The distance $d(v,u)$ affects whether destination node u receives false of legitimate route. So does the risk, and we reasonably assume that risk is proportional to distance:

$$R_v \sim \sum_{i=1} d(v,u);\ \ u \in V \quad . \tag{1}$$

The expression (1) denotes relative quantity of route hijack risk for node *v* regarding target group of network nodes *V*. One cannot predict whether destination node *v* receives false or legitimate route since there are no ways to see the BGP processes inside *v* in real time. But one can make personal probability estimate. Let's call it "trust". The matter of trust is probability that node *v* receives and uses (and further propagates) legitimate route originated by *u*. The value of trust *T* is a ratio of average distance between *v* and other nodes, and the distance between *v* and particular *u*:

$$T_u^v = \frac{\sum_i d(i,v)}{d(u,v)(|V|-1)}\ ;\ \{i,u,v\} \in V\ \ ; u \neq v\ ; u \neq i \tag{2}$$

The risk depends on two components – loss and likelihood, and the latter is very similar to probability. So, we got a new metrics for Internet nodes related to route protection.

If we express likelihood via trust, let's express losses using the number of nodes impacted by false routes due to route hijack. The shorter paths $\pi(p_v)$ go through node *v* or prefixes originating from it, the greater is the impact of this node upon routes distribution. This parameter is calculable by BGP routing tables. Let's call it "significance" S_V^u :

$$S_V^u \sim |\pi_V(p)| \tag{3}$$

Significance should characterize node *v* in terms of number of IP addresses which might potentially use routes received through *v*. It is impossible to know the exact number, so we offer a simplified estimation based on quantity and weight of network prefixes announced via *v*. By "weight" we mean the amount of IP addresses covered by network prefix:

$$w_\pi = 2^{24-l(\pi)} \tag{4}$$

In (3) w_π is the weight of prefix π, $l(\pi)$ is the length of prefix π. Consequently, the weight of the prefix length of 24 bits and covering 256 addresses (which is the least prefix to appear in global routing) is 1, and, for example, weight of 16-bit prefix covering 32768 addresses is 256.

In addition, it should be noted that for each network receiving forged route via *v,* the node *v* also has a certain trust metric. Thus, the degree of influence of the route received from the provider's node will have the greatest impact, because the distance to the provider is the smallest. Taking this into account, we offer to consider each prefix π with relaxing coefficient $(1+\delta)^{-1}$ where *δ* is a metric distance between prefix origin and target node *v*. For example, prefixes originating directly from *v* will have *δ*=0, and will be considered with coefficient $(1+\delta)^{-1}$ =1. Thus, the significance metrics will have the following form:

$$\sum_{\pi} \qquad \sum_{\pi} \tag{5}$$

where δ_π is metric distance between prefix origin and node *v*.

Two metrics create model risk-oriented node distribution in a 3-dimentional space (*R,T,S*):

$$R_v^u = T_v^u S_v^u \tag{6}$$

where R_v^u is the risk of hijacking routes originated by *u* on particular node *v*, T_v^u is a trust metrics of *v* evaluated by *u*, S_v^u is significance metrics of *v* evaluated by *u*. And there is an integral risk R^u of route hijacks through a set of Internet nodes *V*:

$$R^u = \sum_{i \neq u} R_i^u \tag{7}$$

7 Empirical Study: Risk Assessment and Mitigation

For experiment we processed real BGP routing tables of several autonomous systems, and got the real node risk distributions. Here we analyze AS8258, AS6939 (Hurricane Electrics), AS15645 (Ukrainian Internet Exchange). For each study we found the most significant nodes and measured trust to each of them from the viewpoint of AS8258. Figure 5 represents nodes ordered by descending significance metrics. Figure 6 represents trust metrics for nodes in the same order as before. Then, we calculated the risk of route hijack for each node according to (6) and integral risk according to (7).

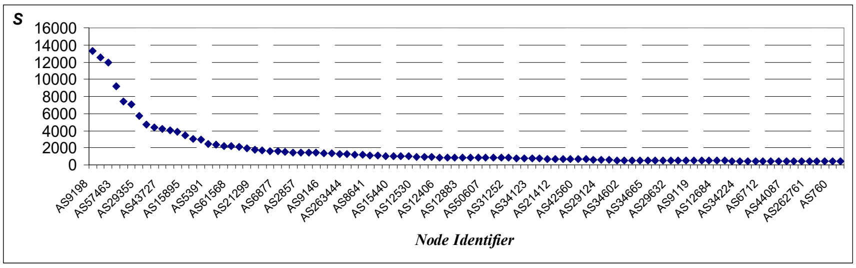

Fig. 5. Node distribution by significance metric *S* for initial topology.

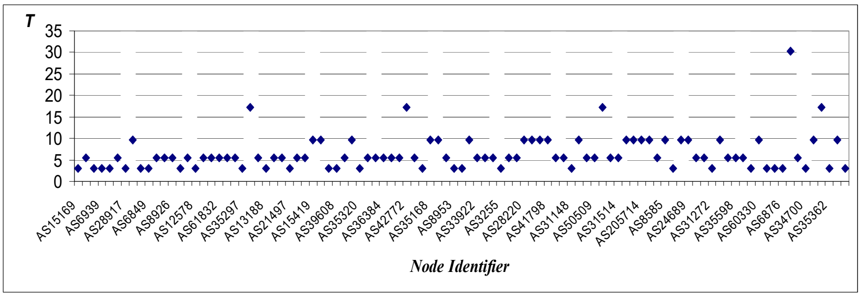

Fig. 6. Trust metric for nodes ordered by significance.

Figure7 represents nodes (AS identifiers) ordered by risk. The integral risk of this model is $R^u = 1098206$.

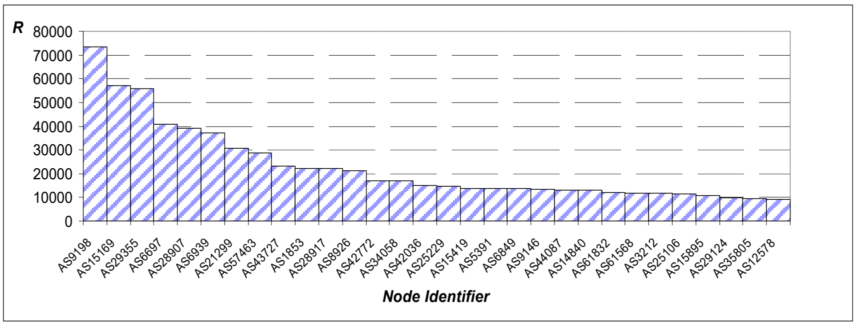

Fig. 7. Node distribution by risk for initial topology.

Then we modified source routing data pretending that AS8258 has direct links to 3 most risky nodes, recalculated metrics and risk, and received a result on Figure 8.

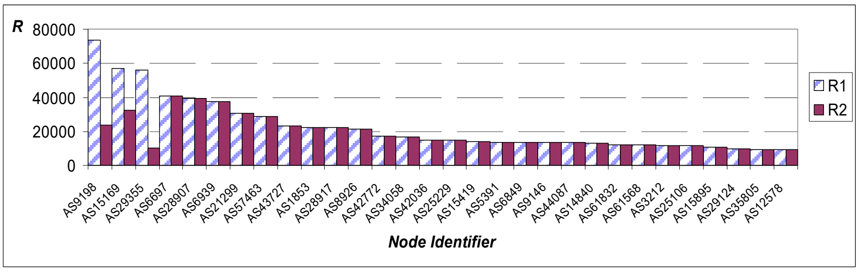

Fig. 8. Comparison of risk for resulting topology (R2) to initial topology (R1).

The integral risk of this model is $R^{u} = 950756$. So, by modeling the new topology using trust metric, we have achieved the risk reduction of approximately 15%.

8 Conclusion

The most significant problems deriving from Border Gateway Protocol weaknesses and vulnerabilities are route leak and route hijack threats. An important step towards assessing the risk posed by attacks on global routing is to predict the impact of the attack, namely to assess the scale of the attack (distribution routes, impact area, number of "damaged" routes). Estimating the risks of route hijack requires quantitative measurement of the impact of an attack on the routing distortion, and therefore, the loss of information through security breach.

There is the relationship between the topology of the Internet and routing vulnerability. We formulated and proposed an approach for assessing and mitigation of route hijack risk using two new metrics for Internet nodes derived from topology learning – trust and significance. While the significance metric describes the evaluation of potential losses in case of hijacking target route on a particular node, the trust metric helps us evaluate the likelihood of route hijack on particular node. Both metrics together are two components of information security risk related to attacks on global routing.

Empirical studies confirm the hypothetic assumption, that measuring the risk opens the way for developing ways of improvement of AS links topology towards higher information security by mitigating the possible risks of attacks on global Internet routing.

References

1. “Internet Mapping and Annotation. Center for Applied Internet Data Analysis” [Online]. Available: https://www.caida.org/research/topology/internet_mapping/. Accessed on: June 28, 2020.
2. Newman M. “The structure and function of complex networks”. *SIAM Review,* 2003,Vol.45:167–256.
3. Faloutsos M.,Faloutsos P., and Faloutsos C. “On Power Law Relationships of the Internet Topology”, *Comput. Commun. Rev.*,1999, №29:251-263.

4. Zubok, V. “Retrospective Analysis Cyber Incidents Related to Attacs on Global Routing”. *Modelyuvannya ta informaciyni technologii(Modeling and Information Technologies)*: Coll. of Scientific Papers. 2019,№86:41-49.DOI:10.5281/zenodo.3610642.
5. “RIPE NCC's Implementation of Resource Public Key Infrastructure (RPKI) Certificate Tree Validation” [Online]. Available: https://tools.ietf.org/html/rfc8488. Accessed on: May 25, 2020.
6. Zubok,V. “Metric Approach to Risk Evaluation of Cyberattacks on Global Routing” : *Selected Papers of the XVIII International Scientific and Practical Conference "Information Technologies and Security" (ITS 2018)* : Vol-2318 urn:nbn:de:0074-2318-4.
7. P. Sermpezis, and V. Kotronis, et al. “ARTEMIS: Neutralizing BGP Hijacking within a Minute” arXiv:1801.01085v4 [cs.NI] 27 Jun 2018.
8. T. McDaniel, J.M. Smith, and M. Schuchard. “Peerlock: Flexsealing BGP”, arXiv:2006.06576v3 [cs.NI] 17 Jul 2020.
9. “Risk Management – Vocabulary (ISO Guide 73:2009, IDT) : DSTU ISO Guide 73:2013. – [Validsince 2014–07–01] .”, Kyiv : Minekonomrozvytku Ukrainy : 2014.
10. Y. Rekhter, P. Gross. “Application of the Border Gateway Protocol in the Internet (RFC 1772)” [Online]. Available: http://tools.ietf.org/html/rfc1772. Accessed on Sep 20, 2019.
11. V.Zubok. “Building Formal Model of the Internet Routing for Risk Evaluation of Cyberattacks on Global Routing”. *CEUR workshop Processing* : 2020: Vol.2577:292-301. [Online]. Avaliable: http://ceur-ws.org/Vol-2577/. Accessed on Aug 12, 2020.