

Evaluating and Improving Regional Network Robustness from an AS TOPO Perspective

Yujia Liu*, Changqing An*, Tao Yu*[†], Zhiyan Zheng*, Zidong Pei*, Jilong Wang*[†], Chalermpol Charnsripinyo[‡]

*Institute of Network Sciences and Cyberspace, BNRist, Tsinghua University, Beijing, China

[†]National Electronics and Computer Technology Center,

National Science and Technology Development Agency, Pathum Thani 12120, Thailand

Email: liuyujia19@tsinghua.org.cn, {acq,zhzy,wjl}@cernet.edu.cn,

{yu_tao,pei zidong}@tsinghua.edu.cn, chalerm.pol.charnsripinyo@nectec.or.th

Abstract—Currently, regional networks are subject to various security attacks and threats, which can cause the network to fail. This paper borrows the quantitative ranking idea from the fields of statistics and proposes a ranking method for evaluating regional resilience. Large-scale simulated failure events based on probabilistic sampling is performed, and a significance tester that measures the impact of events from the overall level and variance aspect is also implemented. To improve a region’s robustness, this paper proposes a greedy algorithm to optimize the resilience of regions by adding key links among AS. This paper selects the AS topology of 50 countries/regions for research and ranking, evaluating the topology robustness from connectivity, user, and domain influence perspectives, clustering the results and get typical region types, and adding optimal links to improve the network resilience. Experimental results illustrate that the resilience of regional networks can be greatly improved by establishing a few new connections, which demonstrates the effectiveness of the optimization method.

Index Terms—Autonomous System (AS), network resilience, network measurement

I. INTRODUCTION

The Internet has become one of the key infrastructures on which all aspects of people’s lives depend. As the basis for ensuring stable Internet communication, network availability is critical. The network of a country or region is subject to various security attacks and threats. Various types of malicious people, such as hackers and terrorists, are attempting to find network topology weaknesses to disrupt or intercept Internet traffic [1]. If cyberspace security is analogized to weapons and tactics, then macroscopic network topology is the key terrain of cyberspace. Attackers targeting weak components can disrupt, manipulate or eavesdrop on internal or international communications of a country or region.

There is currently no work to comprehensively assess the resilience of each region and address the following questions: Is there any difference in the resilience of each region, and if so, how big is the difference; what is the key weak topology that causes such a gap; and how should the region optimize the topology to improve its own resilience? We conducted a comprehensive assessment of the resilience of regional networks to solve the above problems, and make four major contributions.

Assess the resilience of each region and find the key weak components: To address these problems, we propose a statisti-

cal method to evaluate the resilience of a region under attack. We simulate failure event according to the probability of the event to approximate the damage caused by the simulated event in the real situation. For a comparative analysis of regional resilience, we implement a significance tester using the Kruskal-Wallis test [21] and Levene’s test [26] on the resilience samples to rank them at the overall level and the variance level, and finally get ranking of 50 regions. We cluster the regional resilience at the overall level and variance aspect and get several typical types of invulnerability.

Optimize the topology of each region: After finding the key weak components, we propose an optimization objective formula for improving regional resilience and an algorithm based on greedy search. The optimal AS links that should be added for fifty regions to improve intra-region network topology are rendered. Also, we give the optimal suggestion for the boundary AS connection to improve inter-region resilience. Experiments illustrate that the proposed algorithm would improve the resilience of the regions to a large extent while controlling the cost of establishing connections.

Construct an AS topology with region labels: Based on the measurement data obtained from open measurement platforms, we propose a voting-based IP geolocation method and a lightweight AS geolocation method and construct an AS topology with region labels.

Resource weight and influence are considered: We use resource weights to evaluate the impact of a failure and resource weights are evaluated from connectivity perspective, user influence, and domain influence perspective.

To facilitate future research, we make our code and data from our experiments available in [27].

This paper is organized as follows. Section 2 introduces existing research efforts, section 3 presents ranking method for evaluating regional resilience, section 4 proposes a greedy algorithm to optimize the resilience of regions, section 5 conducts a thorough evaluation, and section 6 concludes the paper.

II. RELATED WORK

Given the importance of Internet connectivity, researchers have conducted extensive research on its resilience, dependencies, and key components.

In terms of Internet resilience, Wu et al. [2] proposed a failure model that systematically analyzed the impact of various failures on Internet communication using network reachability and network traffic-based metrics. Dolev et al. [3] analyzed the resilience of the Internet AS topology under failures or attacks based on the business agreement between ASes. They found that the Internet became less resilient in the aspect of deliberate attacks and slightly less resilient under random failures. They used different metrics to measure internet resilience and analyze the impact of backup connections on internet resilience. Luckie et al. [4] studied the dependency of the Internet on a single router at the AS level and proposed an effective and active detection technique. Experiments showed that the proposed active detection technology proposed by Luckie et al. [4] was able to detect router restarts in real-time with high accuracy.

In terms of the survivability of key components under attacks, Omer et al. [5] proposed a model for measuring the resilience of submarine optical cable networks, defining the transmission value before and after network outages to measure the basic resilience of the network. CAIDA project MapKit [6] proposed to find the main components of the national or regional Internet topologies, including autonomous systems, Internet Exchange Points (IXPs), and regional Points of Presence (PoPs). This project mapped the relationship between the components and studied the sensitivity of the national or regional Internet topology to connection interruptions.

In terms of routing dependencies, Karlin et al. [7] developed the first framework for country-level routing analysis, focusing on the impact of individual countries on international routing. The findings showed that the United States, the United Kingdom, and Germany were the core of international routing, whose policies have a greater impact on international routing. Leyba et al. [8] defined the National Chokepoint Potential (NCP), which was used to evaluate the influence of the boundary AS in the regional topology and compared 8 regions horizontally to analyze the changes and differences of the NCP. Gamero et al. [9] studied the influence of transmission operators at the regional level and found out the regions in which traffic were less likely to be observed and manipulated by transmission operators in other regions and have higher security. Edmundson et al. [10] studied the extent to which a country had to depend on the United States and other western countries for their routing paths. Acharya et al. [11] studied the phenomenon of routing centralization and found the key players that support the Internet backbone.

Among them, Wu et al. [2], Dolev et al. [3], Luckie et al. [4], Omer et al. [5] did not study the regional granularity. The CAIDA project MapKit [6] studies the resilience of the region, but does not focus on the elasticity differences before the region. Karlin et al. [7], Lebba et al. [8], Gamero et al. [9], Edmundson et al. [10], Acharya et al. [11] focus on dependence on certain nodes, not an assessment of the overall resilience of the region.

III. REGIONAL SURVIVABILITY

A region consists of multiple ASes. In this section, we will first define AS survivability, and then define regional survivability based on ASes.

A. AS Survivability

The remaining connections between an AS and other nodes after a failure event occurs is called the survivability of the AS to a failure event. The survivability of AS i to an event x is defined as:

$$e_{ix} = \frac{\sum_{b \in B} r_b}{\sum_{o \in O} r_o}. \quad (1)$$

The denominator is the sum of all AS node resources in the region, and the numerator is the sum of AS node resources that still can connect to the AS node i after the event. The region can be the Internet, a country, or a region.

O represents the set of regional nodes, B represents the set of nodes that can communicate with the node i after the event, and r represents the resources owned by a node, such as the number of users of a node or the important domains of a node.

By simulating failure events multiple times based on probability, the survivability samples $E_i = [e_{i1}, e_{i2}, \dots, e_{i(m-1)}, e_{im}]$ of the node i can be obtained, where m is the total number of samples of node i . It is a one-dimensional vector called the survivability sampling of an AS. Through hierarchical probability sampling, this paper simulated damage events according to probability of the events to approximate the survivability with regards to real-life damage events.

B. Resource Weight

Resource weights are used to evaluate the importance of different nodes in the network. For the i th AS node AS_i , the resource weight r_i is defined from three perspectives:

- From a connectivity perspective. Each AS has the same weight, so r_i is 1. The concern is ASes that are affected.
- From the perspective of user influence. r_i is the user proportion of AS_i . We use the proportion of users of every AS in the whole region as the weight from the perspective of user influence and calculate it based on [12]. The concern is the affected users.
- From the perspective of domain name influence. r_i is the domain importance measure of AS_i . We generated domain weight based on the results of [13], applied Zipf distribution to assign weights to each prefix, and aggregated prefix weight into AS. The concern is the affected resource.

C. Routing Influence

Routing influence is used to measure the influence of the routing service provided by an AS node for other ASes in a certain region, indicating the percentage of routing paths passing through the AS node.

The influence of AS_b to AS_o is defined as:

$$R_i = \frac{P(AS_b, AS_o)}{r(AS_o)} \quad (2)$$

where $P(AS_b, AS_o)$ is the number of routings of AS_o that go through AS_b and $r(AS_o)$ is the number of routings.

However, measurement data collection is heavily biased towards the data seen by probe points. The limitations and biases of the detection point itself had a great influence on the result [14]. In order to measure the actual influence of ASes in a region, a filter is designed to filter out the detection points that are too far from or too close to the specific AS in order to mitigate the influences from limitations and unbalanced distributions of detection points.

The routing influence $w_c(AS_b)$ of AS_b to a region is defined as:

$$w_c(AS_b) = \sum_{o \in O} \theta_o \times UR_c(AS_b, AS_o) \quad (3)$$

where O indicates all ASes in the region c , $UR_c(AS_b, AS_o)$ indicates the routing influence on AS_o based on user weights, and θ_o indicates the % of users AS_o in region c . The routing influence based on the user weight UR is defined as:

$$UR_c(AS_b, AS_o) = \frac{1}{n - 2\alpha n} \sum_{i=\alpha n+1}^{i=\alpha n+1} R[i] \quad (4)$$

where R is a one-dimensional vector that stores the ratio of the actual observed AS_o passing AS_b at all detection points. α is used to filter the detection points whose distances are too close or too far and is set as $\alpha = 0.1$.

D. Ranking Regional Survivability

Regional survivability is defined as the weighted sum of its AS survivability. Fig. 1 depicts the calculation process of regional survivability.

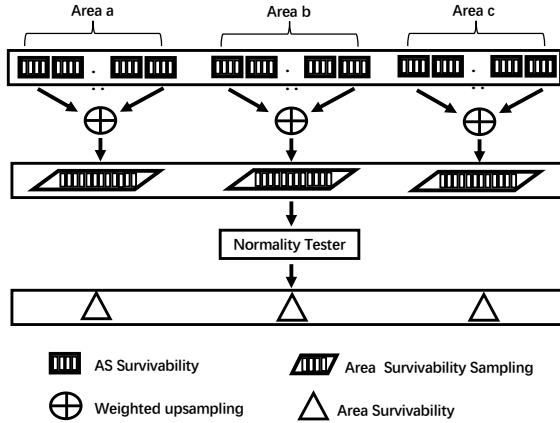


Fig. 1. Calculation process of region survivability

First, for each region, select a set of significant ASes and calculate the survivability sampling for each AS. Second, sample from each AS samples by routing influence. Finally, for the sampled data of each region, Kolmogorov-Smirnov [25] is used for the normality test, and Hartley [22] is used for the homogeneity of variance test. It is found that the sampled data satisfied the normal distribution but not the homogeneity of variance. As a result, Kruskal-Wallis test [21] method

is chosen to compare two regions at the overall level, and Levene's test [26] method is chosen to compare two regions at the variance level. The function $rank_r()$ is created to output the ranking result.

The complexity of computing connectivity between ASes is reduced by building a routing tree [15]. We calculate the connectivity change before and after a failure event based on the routing tree. In the worst case, the time complexity is $O(m + n)$, where m is the number of AS connections, and n is the number of ASes.

For each region, intra-region survivability and inter-region survivability are calculated respectively. Intra-region survivability focuses on the connectivity among ASes geographically located in a region.

Inter-region focuses on the connectivity between a region and the other regions. We select all the boundary ASes of a region as important nodes and calculate the survivability sampling of the boundary AS.

In the calculation of inter-region survivability, for the region c_i , the resource weight r_i of a node is defined from three perspectives:

- From the perspective of connectivity, r_i is always 1.
- From the perspective of regional economic influence, r_i is the Zipf weight of regional c_i GDP per capita.
- From the perspective of the importance of domain names, r_i indicates the importance of the domain names in the region c_i .

The connectivity weight focuses on the number of connected regions and the regional economic influence is measured by the regional per capita GDP. It is known that a large number of events on the Internet follow the Zipf distribution [16]. To prevent the gap between the regional GDP per capita from being too large, the Zipf distribution is used to reprocess the per capita GDP data, and the obtained Zipf weight is used as the regional economic weight r_i . The domain name weight of a region is calculated as the sum domain name weights of all ASes within the region.

E. Vulnerable Groups

If the failure of a group of nodes will result in many connection failures, these nodes constitute a vulnerable group. When we simulate failures, we note down the loss of connection due to the event and rank them to get vulnerable groups of a region.

IV. OPTIMIZING REGIONAL SURVIVABILITY

A. Objective

The goal of optimization is to improve the survivability of the region. We implement this by adding some AS connections under certain cost limitations.

The target formula of the optimization algorithm is:

$$\begin{aligned} \min_Q \quad & rank_r(topo_{old}, Q) \\ \text{s.t.} \quad & cost(Q) < c \end{aligned} \quad (5)$$

where $rank_r()$ outputs the ranked value of the optimized region r , $topo_{old}$ represents the original topology of the region r , Q is the set of optimal connections, and c is the cost limit to establish the optimal connection Q , which can be a fixed value defined by the administrator.

For each vulnerable group, we calculate the node pairs that cannot communicate and store the source node and destination node in the form of $L = AS_W : ([n_i, n_j], \dots, [n_k, n_l]), \dots$, which indicates that after the weak group AS_W is destroyed, n_i and n_j cannot communicate, and n_k and n_l cannot communicate.

The optimization objective can be transformed to maximizing the utility of the reconnected node pairs of the set L , and then we get:

$$\begin{aligned} \max_Q \quad & re_value_r(topo_{old}, Q) \\ s.t. \quad & cost(Q) < c \end{aligned} \quad (6)$$

where $re_value()$ represents the utility of the reconnected node pair after establishing the optimal connections Q .

We can remove the constraints and introduce a parameter α to control the relative value of the utility and the cost to reduce the difficulty of designing the solution algorithm.

$$\max_Q \quad re_value_r(topo_{old}, Q) - \alpha \times cost(Q) \quad (7)$$

B. Utility

Utilities are the resources weight of the nodes that can communicate after establishing the optimal connection. For node pairs with broken connectivity $L = AS_W : ([n_i, n_j], \dots, [n_k, n_l])$, if after establishing optimized connection q_i , nodes n_i and n_j can re-communicate with nodes n_k and n_l , then the benefit of this optimized connection is the sum of the resource weights of the nodes n_i, n_j, n_k and n_l .

C. Cost

Different commercial connections, such as P2P (Peer to Peer) connection or a P2C/C2P (Provider to Customer/Customer to Provider), can be established between ASes. ASes typically establish P2C connections with lower-tier ASes and set up P2P connections with tier equivalence ASes. In a P2C relationship, the customer pays the provider, while in a P2P relationship both parties do not need to pay to other. In subsequent experiments, the cost of P2C relationship is set higher than that of P2P.

A P2P relationship can be established when a pair of ASes $\langle AS_p, AS_q \rangle$ satisfies one of the following conditions: AS_p and AS_q are at the same level in the network; AS_p and AS_q have a relatively different level in the network, but AS_p has established a P2P relationship with an AS equivalent to AS_q in terms of their levels and vice versa. Otherwise, $\langle AS_p, AS_q \rangle$ will establish a P2C/C2P relationship.

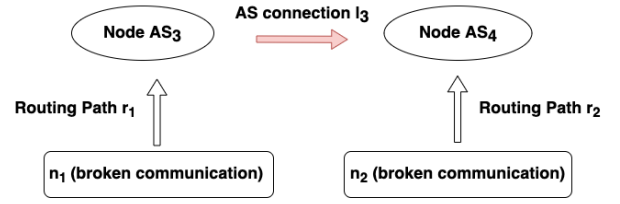
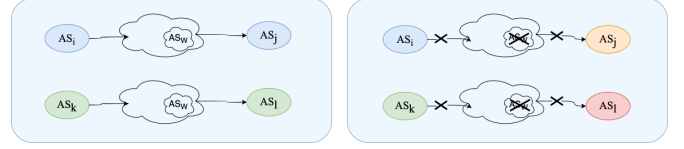
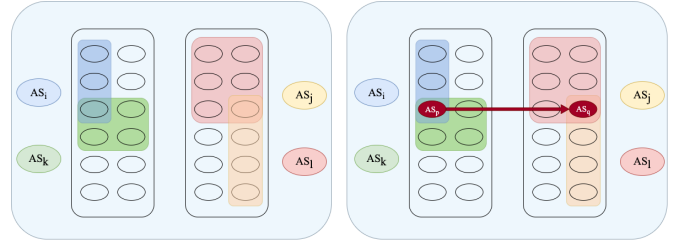


Fig. 2. The AS relationship and link optimization



(a) calculating the node pairs that can't communicate



(b) greedy search

Fig. 3. Searching for the optimal link

D. Method

Fig. 2 shows potential optimized connections for two node pair $[n_1, n_2]$. There is no connection between AS_3 and AS_4 before optimization, and n_1 communicates with AS_3 through the routing path relationship r_1 , and n_2 communicates with AS_4 through the routing path relationship r_2 . It is expected to search for a pair of ASes $\langle AS_p, AS_q \rangle$ which can make $[n_1, n_2]$ communication. From the figure we can see if AS_3 and AS_4 can successfully establish the l_3 connection, then n_1 and n_2 may communicate successfully.

Considering the valley-free principle [2], r_1, r_2, l_3 must meet the following conditions.

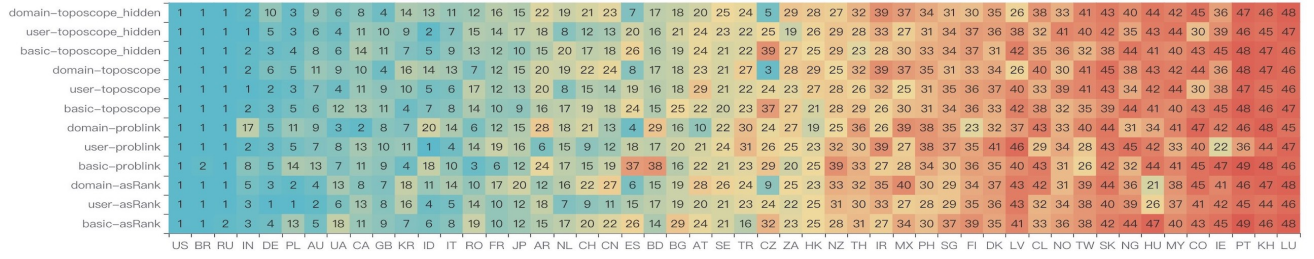
First, let us define the routing path type:

- $s_1 : c2p[n]$,
- $s_2 : c2p[0/n] \ \& \ p2p[0/1] \ \& \ p2c[0/n]$.

where $n > 1$. $r[n]$ means there are n consecutive connections with the r relationship in the routing path, $r[0/n]$ means there exists 0 or n consecutive connections with the r relationship in the routing path, $r[0/1]$ means there exists 0 or 1 connection with the r relationship in the routing path, and the symbol $\&$ indicates that $c2p[0/n]$, $p2p[0/1]$, and $p2c[0/n]$ are adjacent in the routing path.

Considering the valley-free principle [2], the following form of routing path relationship will not occur: $p2c[1/n] \ \& \ p2p[0/1/n] \ \& \ c2p[1/n]$, where $n > 1$.

Table I shows all possible matching relationships, and it is necessary to calculate the optimal connection according to matching relationships.



(a) overall level



(b) variance level

Fig. 4. Ranking of Intra-region survivability

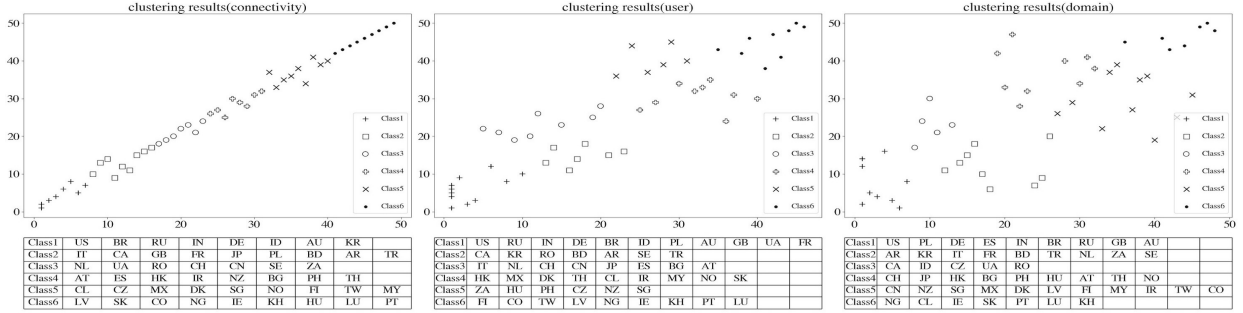


Fig. 5. Clustering of intra-region resilience

TABLE I
RELATIONSHIP AND POSSIBLE CONNECTION

	Relationship Between n_1 and AS_3	Relationship Between n_2 and AS_4	Possible Connection Between AS_3 and AS_4
Case1	s_1	s_1	p2p/p2c/c2p
Case2	s_1	s_2	c2p
Case3	s_2	s_1	p2c

It can be found that s_1 relationship has higher adaptability by matching with both s_1 and s_2 relationships, and the cost to establish the optimal connection is more likely lower. When two s_1 relationships match, an optimal connection of P2P relationship can be established. The cost of P2P relationship is lower than the cost of C2P/P2C relationships.

Fig. 3 shows the searching process for potential optimal connections.

Based on the routing tree of each node, we compare the nodes on the routing tree before and after the weak group is

destroyed, and obtain the node pairs that cannot communicate after the weak group is destroyed, as shown in Fig. 3(a). The weak group AS_W may consist of multiple AS nodes and links. When nodes and links in AS_W are destroyed, AS_i and AS_j can't communicate, neither can AS_k and AS_l . We store pairs of nodes that cannot communicate as a Bipartite graph.

Fig. 3(b) shows the idea of greedy search. After the node pairs that cannot communicate have been calculated, we search the routing tree for each node in the pairs, to find the nodes in the network each node can communicate with. Then we search for the node that can communicate with most of the left nodes and right nodes respectively.

As shown in Fig. 3(b), the nodes with the same background color can communicate with each other. Among them, AS_i , AS_k can communicate with AS_p , and AS_j , AS_l can communicate with AS_q under certain match rules. The greedy algorithm is used to search for the optimal connection with the highest utility. In Fig. 3(b), AS_p , and AS_q are the nodes with highest utility, and we can establish an optimal connection between them.

V. EVALUATION

A. Data

We used 4 TOPO data sets (ASRank [18] in August 2022, Problink [19] in 2021, Toposcope [20] in 2019 and h-Toposcope which combined Toposcope data and the inferred hidden connections) and 3 resource weight combinations for evaluation, and obtained a total of 12 evaluation results together. We selected 50 regions to do the experiment. We used the number of users data provided by APNIC as the user weight [12]. We generated domain weight based on the results of [13], applied Zipf distribution to assign weights to each prefix, and aggregated prefix weight into AS.

The AS may be located in multiple regions. Although the Regional Internet Registry (RIR) registers the region to which each AS belongs when assigning the AS number, it only registers the region to which each AS belongs in the legal sense. We utilized BGP routing data from Routeviews [24], Traceroute data from RIPE ATLAS [23] and IP location data in 2022. We propose a fast scheme for AS geolocation by combining the IP location database, BGP routing data and Traceroute data. First, we locate the IP address to the AS and the region by searching the IP geolocation database and using the IP as the medium to obtain the corresponding relationship between the AS and the region. For ASes located in a single region, the region location process is complemented. We find that 6.8% of ASes are located in multiple regions. For an AS located in multiple regions, we implement an algorithm based on the context of the current AS (the ASes before and after the current AS in the routing path) by referring to the regional path estimation algorithm proposed by Karlin et al. [7].

B. Ranking

Fig. 4(a) and Fig. 4(b) depict the network resilience for 50 regions at the overall level and the variance level, respectively. The regional resilience metric is a relative ranking and it is displayed in a heat map. The smaller the ranked value is, the better the regional resilience is. The redder the color in the heat map represents, the larger the ranked value is. The bluer the color represents, the smaller the ranked value is. The x-axis of the heat map represents 50 regions, and the regions are represented by region codes. The y-axis represents different AS relationship data sets and resources weight combinations.

Fig. 4(a) shows the intra-regional survivability rankings at the overall level. US, BR, RU, IN, DE ranks at the forefront. Fig. 4(b) shows the regional survivability ranking under the variance aspect and there are some differences between the overall and the variance ranking, which shows the rationality of the joint evaluation of the two indicators.

We cluster the regional overall ranking result and the variance ranking result. Fig. 5 shows the clustering results under connectivity weight, user weight, and domain weight. The x-axis is the overall level and the y-axis is the fluctuation level under the four topo data combinations. On connectivity perspective, US, RU, BR perform better with higher survivability at both overall and variance levels. There are some differences when clustering on user and domain perspectives.

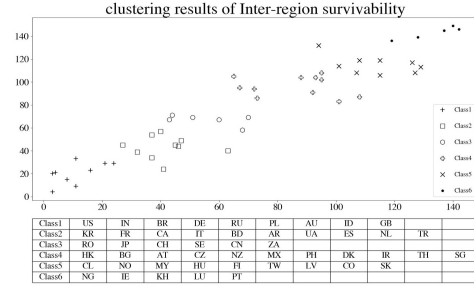


Fig. 6. Clustering of inter-region survivability

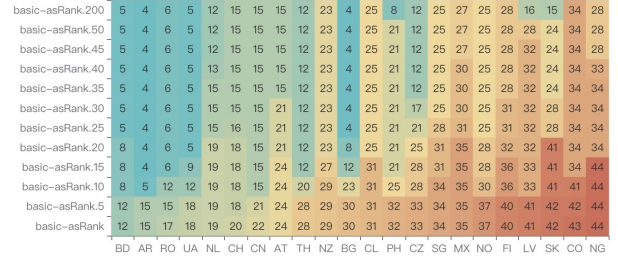


Fig. 7. Intra-region ranking change after optimization

Fig. 6 shows the clustering results of the inter-region resilience from the overall level and the fluctuation level.

C. Optimizing intra-domain resilience

Fig. 7 plots the ranking of resilience for regions with optimal links. Numbers on the y-axis indicate the number of optimal connections. The Experiment showed that current optimization algorithm can improve the resilience of the region to a large extent. There is still more work that needs to be done.

D. Optimizing inter-region resilience

The ranked value of inter-region resilience is proportional to the ranking of the regional border ASes. To optimize inter-region resilience, increasing the number and ranking of regional border ASes is an effective measure, such as having border ASes connected with multiple regions.

VI. CONCLUSION AND FUTURE WORK

This paper borrows the quantitative ranking idea from the fields of statistics and proposes a ranking method for evaluating regional resilience. Then, it selects vulnerable groups from different regions, formulates an optimization goal, and proposes a greedy optimization scheme. More work could be done in the future to further enhance regional resilience, including establishing PoPs, enriching physical facilities, and other related initiatives.

ACKNOWLEDGMENT

This work is supported by a) the National Key Research and Development Program of China (No.2020YFE0200500), and b) the APNIC ISIF Funding “Developing a Collaborative BGP Routing Analyzing and Diagnosing Platform”.

REFERENCES

- [1] A. Kozłowski, "Comparative analysis of cyberattacks on Estonia, Georgia and Kyrgyzstan," *European Scientific Journal*, vol. Vol 3, pp. 237–245, 11 2020.
- [2] J. Wu, Y. Zhang, Z. M. Mao, and K. G. Shin, "Internet routing resilience to failures: analysis and implications," in *Proceedings of the 2007 ACM CoNEXT conference*, 2007, pp. 1–12.
- [3] D. Dolev, S. Jamin, O. O. Mokryn, and Y. Shavitt, "Internet resiliency to attacks and failures under BGP policy routing," *Computer Networks*, vol. 50, no. 16, pp. 3183–3196, 2006.
- [4] M. Luckie and R. Beverly, "The impact of router outages on the AS-level internet," in *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, 2017, pp. 488–501.
- [5] M. Omer, R. Nilchiani, and A. Mostashari, "Measuring the resilience of the global internet infrastructure system," in *2009 3rd Annual IEEE Systems Conference*, 2009, pp. 156–162.
- [6] CAIDA, Mapkit - investigating the susceptibility of the internet topology to country-level connectivity disruption and manipulation, 2017. [Online]. Available: <https://www.caida.org/funding/satc-mapkit/>
- [7] J. Karlin, S. Forrest, and J. Rexford, "Nation-state routing: Censorship, wiretapping, and BGP," *arXiv preprint arXiv:0903.3218*, 2009.
- [8] K. G. Leyba, B. Edwards, C. Freeman, J. R. Crandall, and S. Forrest, "Borders and gateways: measuring and analyzing national as choke-points," in *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, 2019, pp. 184–194.
- [9] A. Gamero-Garrido, E. Carisimo, S. Hao, B. Huffaker, and K.C. Claffy, "Inferring country-level transit influence of autonomous systems, 2020. [Online]. Available: https://www.caida.org/publications/presentations/2020/inferring_country-level_transit_influence_kismet/.
- [10] A. Edmundson, R. Ensafi, N. Feamster, and J. Rexford, "Nation-state hegemony in internet routing," in *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, Jun. 2018, pp.1-11.
- [11] H. B. Acharya, S. Chakravarty, and D. Gosain, "Few throats to choke: On the current structure of the internet," in *2017 IEEE 42nd conference on local computer networks (LCN)*, 2017, pp.339–346.
- [12] APNIC, Visible ASNs: Customer populations (est.), 2017. [Online]. Available: <https://stats.labs.apnic.net/aspop/>.
- [13] J. Naab, P. Sattler, J. Jelten, O. Gasser, and G. Carle, "Prefix top lists: Gaining insights with prefixes from domain-based top lists on DNS deployment," in *Proceedings of the Internet Measurement Conference*, 2019, pp.351–357.
- [14] R. Fontugne, A. Shah, and E. Aben, "The (thin) bridges of as connectivity: Measuring dependency using AS hegemony," in *Passive and Active Measurement*. Springer International Publishing, 2018, pp.216–227.
- [15] P. Gill, M. Schapira, and S. Goldberg, "Modeling on quicksand: Dealing with the scarcity of ground truth in interdomain routing data," *ACM SIGCOMM Computer Communication Review*, vol. 42, no. 1, pp. 40–46, Jan. 2012.
- [16] S. A. Krashakov, A. B. Teslyuk, and L. N. Shchur, "On the universality of rank distributions of website popularity," *Computer Networks*, vol. 50, no. 11, pp.1769–1780, 2006.
- [17] L. Gao, "On inferring autonomous system relationships in the internet," *IEEE/ACM Transactions on Networking*, vol. 9, no. 6, pp.733–745, 2001.
- [18] X. Dimitropoulos, D. Krioukov, B. Huffaker, k. claffy, and G. Riley, "Inferring AS relationships: Dead end or lively beginning?" in *Experimental and Efficient Algorithms*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp.113–125.
- [19] Y. Jin, C. Scott, A. Dhamdhere, V. Giotsas, A. Krishnamurthy, and S. Shenker, "Stable and practical AS relationship inference with ProbLink," in *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*, Boston, MA, Feb. 2019, pp. 581–598.
- [20] Z. Jin, X. Shi, Y. Yang, X. Yin, Z. Wang, and J. Wu, "Toposcope: Recover as relationships from fragmentary observations," in *Proceedings of the ACM Internet Measurement Conference*, Oct. 2020, pp.266–280.
- [21] N. Breslow, "A generalized kruskal-wallis test for comparing k samples subject to unequal patterns of censorship," *Biometrika*, vol. 57, no. 3, pp.579–594, 1970.
- [22] K. V. Ramachandran, "On the tukey test for the equality of means and the hartley test for the equality of variances," *The Annals of Mathematical Statistics*, pp. 825–831, 1956.
- [23] RIPE Atlas. [Online]. Available: <https://atlas.ripe.net/>.
- [24] University of Oregon Route Views Project. [Online]. Available: <https://www.routeviews.org/routeviews/>.
- [25] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
- [26] H. Levene, "Robust tests for equality of variances," in *Contributions to Probability and Statistics ed. I. Olkin*, 278–292. Palo Alto, CA: Stanford University Press, pp. 278–292, 1960.
- [27] Evaluating and Improving Regional Network Robustness. [Online]. Available: <https://github.com/thudragonlab/Resilience>.