World Wide Web (2015) 18:749–766
DOI 10.1007/s11280-014-0275-3

K-core-based attack to the internet: Is it more malicious than degree-based attack?

Jichang Zhao · Junjie Wu · Mingming Chen · Zhiwen Fang · Xu Zhang · Ke Xu

Received: 5 May 2013 / Revised: 16 October 2013 /
Accepted: 10 December 2013 / Published online: 10 April 2014

Abstract *K-core* (k-shell) is an interesting measure that discriminates the core and fringe nodes in a complex network. Recent studies have revealed that some nodes of high k-core values may play a vital role in information diffusion. As a result, one may expect that attacking the nodes of high k-core values preferentially will collapse the Internet easily. To our surprise, however, the experiments on two Internet AS-level topologies show that: Although a *k-core-based attack* is feasible in reality, it is actually less effective than the classic *degree-based attack*. Indeed, as indicated by the measure *normalized susceptibility*, we need to remove 2 % to 3 % more nodes in a k-core-based attack in order to collapse the networks. Further investigation on the nodes in a same shell discloses that these nodes often have drastically varying degrees, among which are the nodes of high k-core values but low degrees. These nodes cannot lead to sufficient link deletions in the early stage of a k-core-based attack, and therefore make it less malicious than a degree-based attack. Finally, a strategy called "ELL" is employed for the Internet enhancement. Experiments demonstrate that "ELL" can greatly improve the Internet robustness at very small costs.

Keywords Robustness · K-core index · Malicious attack · AS-level internet

J. Zhao · M. Chen · Z. Fang · K. Xu (✉)
State Key Laboratory of Software Development Environment, Beihang University, Beijing, China
e-mail: kexu@nlsde.buaa.edu.cn

J. Wu
Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, School of Economics and Management, Beihang University, Beijing, China

X. Zhang (✉)
National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing, China
e-mail: zhangxu_cncert@163.com

1 Introduction

The Internet has become the most important communication infrastructure in the world [43], especially after the boom of online social networking sites [28, 29]. Tremendous research efforts have been devoted to scale-free networks, such as the AS-level Internet in the level of autonomous system [3, 6, 8, 36, 37, 46, 49]. Among them, attack survivability is one of the core topics. People find that, while the Internet is robust to the random failure, it is fragile to malicious attacks, which are generally defined as removing important nodes or links preferentially from the networks [1]. Specifically, the simple *degree-based attack*, i.e., attacking the nodes with higher degrees preferentially, is often regarded as the most feasible attack type in reality. For example in 2006, an *Internet Service Provider* called Con Edison(AS number is 27506) announced a number of prefixes owned by other ASes and "stole" Internet traffic from these ASes [12]. In 2009, a misconfiguration triggered a bug in the Cisco *Border Gateway Protocol* implementation, and caused a ten-fold increase in global routing instability for about an hour. This failure was just imported by a small Czesh provider(AS number is 47868) [38]. Other types of attacks, e.g., attacking the nodes with higher betweenness preferentially, may be more malicious than the degree-based attack. But they often need the global topological information of the networks and are very time-consuming [4], and thus become infeasible in practice [19]. Recently, some studies attempt to use local centrality to capture the global betweenness information [16, 27], but its effectiveness for large-scale technical networks (like the Internet) remains unclear.

K-core (k-shell) is an interesting measure that categorizes the nodes in a complex network into the core nodes and the fringe ones. Recently, in their landmark paper [22], the authors found in many types of complex networks that k-core is a more effective measure to describe the influence of a node to the propagation of information or diseases. Indeed, they disclosed a surprising fact that some nodes with high degrees play a trivial role in the information spreading process. They argued that a k-core viewpoint is more instructive; that is, those high-degree nodes actually have low k-core index values and thus locate in the fringe of the network. From this aspect, one may expect that a *k-core-based attack*, i.e., attacking high k-core index nodes preferentially, can collapse the Internet more easily than a degree-based attack. This motivates our study on the k-core-based attack, which to our best knowledge is among the first few studies along this line.

To this end, we performed comparative studies on the two types of malicious attacks. Six measures including both the structural and propagative ones were introduced to characterize the damages to the networks during the attacks. To our surprise, the results on two real-world AS-level Internet data sets showed that: Although a k-core-based attack is feasible using the *traceroute* tool [14, 20], it is less malicious than a classic degree-based attack. Indeed, as indicated by the *normalized susceptibility* measure, we need to remove 2 % to 3 % more nodes in a k-core-based attack to make the network collapsed. Further investigation on the nodes in a same shell disclosed that these nodes often have highly varying degrees, among which are the nodes of high k-core values but low degrees. These nodes cannot contribute sufficient link deletions in an early stage of a k-core-based attack, and therefore make it less malicious than a degree-based attack.

Finally, we paid attention to the enhancement of the AS-level Internet by employing a simple strategy called "ELL" proposed in our previous work [48]. In a nutshell, ELL adds new connections between the nodes of lower degrees preferentially. Experiments demonstrated that "ELL" can effectively improve the Internet robustness at very small costs.

The rest of this paper is organized as follows. Section 2 and Section 3 introduce the related work and preliminaries, respectively. In Section 4, we introduce the real-world

experimental data sets. Section 5 models the malicious attacks and describes the feasibility of a k-core-based attack, with the experimental results given in Section 6. The enhancement of the robustness of the AS-level Internet is discussed in Section 7, and we finally conclude our work in Section 8.

2 Related work

Weak attack survivability but strong error tolerance [1] is a dilemma for the complex networks. In recent years, many researchers focus on the robustness analysis and enhancement of complex networks. For instance, Cohen et al. unveiled that the Internet is resilient to random failures [9] but fragile to the intentional attack [10]. Holme et al. proposed four different attacking strategies and found that attacks by recalculating degrees and betweenness centrality are often more harmful than attacks based on the initial network [19]. An optimal model was presented by Tanizawa et al. to generate robust networks against random errors and malicious attacks [40]. Several approaches for network enhancement were also presented in [34, 42, 48]. The absence of an epidemic threshold in computer virus infections on the Internet was found through the analysis on real data [30]. Then targeted immunization schemes [31] and the immunization of random acquaintances of random nodes [11] were presented to reduce the immunization threshold.

K-core has attracted many research interests from different fields in recent years. For example, Seidman studied the cohesion of the social network by presenting k-cores [35]. Wuchty and Almaas applied a core decomposition method to identify the inherent layer structure of the protein interaction network [41]. Garas et al. employed the k-core decomposition method to quantify the spreading power of a node in the global economic network [17]. Meanwhile, as a key metric in complex networks, k-core also attracts a lot of research interests in the scope of the Internet. For example, Carmi et al. used information on the connectivity of network shells to separate the AS-level Internet into three subcomponents [6]. Zhang et al. found that the k-core with larger k is nearly stable over time for the real AS-level Internet [46]. Zhang et al. proposed a model based on k-core decomposition to model the Internet Router-level topology [47]. In the inspirational work [22], Kitsak et al. focused on evaluating the influence of a node in the spread of information or diseases through its k-core index. They reported an unexpected finding that some hub nodes may locate in the periphery of the network.

Despite of the abundant existing research on the network robustness and k-core index, little work has been done to unveil whether the attack based on k-core is more malicious than other types of attacks to the Internet. This indeed motivates our study in this paper.

3 Preliminaries

In this section, we first discuss the feasibility of attacking the AS-level Internet, and then revisit the measures to characterizing the damages of networks caused by malicious attacks.

3.1 Feasibility of attacking the AS-level internet

The network of the AS-level Internet stands for business relationships between different *Internet Service Providers* (*ISP*). Each AS contains one or several prefixes and different ASes communicate with each other through the *Border Gateway Protocol* (*BGP*) [33], in

which the security problem has not been addressed adequately [5]. In what follows, we show it is indeed feasible to attack one AS in the current Internet.

A recent survey by Bulter et al. revisited several attacking methods [5]. For instance, *prefix hijacking* means an AS A can advertise a prefix from the address space belonging to another AS B; then the traffic that should be routed to B would be routed to A falsely. For another, *link cutting attack* can be manifested by either physically attacking a link or employing Denial-of-Service (DoS) attacks.

In addition, there have been quite a few real-world AS-attacking cases in the history of the Internet [5]. In 1997, a misconfigured router maintained by a small ISP in Florida injected incorrect routing information into the global Internet and claimed to have optimal connectivity to all Internet destinations. As a result, most Internet traffics were routed to this ISP, which overwhelmed the misconfigured router and crippled the Internet for nearly two hours [5]. Another example is the notorious attack launched by Pakistan Telecom (AS number is 17557) in 2008, which announced the prefix belonging to YouTube (AS number is 36561) [45] intentionally. Routers around the world then received this announcement, and redirected YouTube traffics to Pakistan [44].

Some natural disasters may also lead to the failure of the AS-level Internet [38]. For instance, the fire caused by a train derail in the Northeast US caused the disruption of the fibre backbone owned by seven major ISPs. Then most traffics were rerouted and the congestion of alternative links led to a noticeable slowdown of the Internet. Another example happened in 2006, where the earthquake near Hengchun, Taiwan crashed the submarine cables that provide Internet connectivity between Asia and North America.

To sum up, attacking an AS in the real-world Internet is indeed feasible. As a result, it is meaningful to discuss attack survivability of the Internet in the AS-level.

3.2 K-core index and basic measures

The Internet can be intuitively modeled as a graph $G(V, E)$ at different levels, where V is the set of interfaces, routers or ASes, and E is the set of links between them. In this paper, we mainly focus on the AS-level Internet, which means a node stands for an AS and a link stands for the connection between its two ends. The number of links of a node is defined as its *degree*. Then the *averaged degree* of a network can be defined as $\langle k \rangle = \frac{2|E|}{|V|}$.

The distribution of the degree of a graph is denoted as $p(k)$. For the AS-level Internet, $p(k)$ typically follows a power-law distribution [21, 26]. The *Heterogeneity* of a network, defined as $H = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$, is often used to characterize the non-uniformity of the degrees. The *clustering coefficient* of a node i characterizes how closely its neighbors are interconnected [32]. It is defined as $C_i = \frac{2|E_i|}{k_i(k_i-1)}$, where E_i is the set of ties between i's neighbors and k_i is the degree of i. For the case of $k_i = 1$, we set $C_i = 0$. The *average clustering coefficient* of a network can then be defined as $C = \frac{\sum_{i \in V} C_i}{|V|}$.

K-core [35] in a graph G is defined as the maximum subgraph G^k, in which each node's degree is at least k [2]. By recursively pruning the least connected nodes, the hierarchical structure of the network can be broken down to the highly connected central part, which is stated as the core of the network [2, 15]. Then the *k-core index*, denoted as k_s, is used to characterize how far a node is from the core of a network. A node i has a k-core value k_s if it is in the k_s-core but not in the $(k_s + 1)$-core. A larger k_s indicates the more closeness of the node to the core. Hereinafter, we interchangeably use K-core and K-core index when there is no confusion.

K-core can be computed through the following steps [6, 22]. First, remove all the nodes with degree $k = 1$. After this step, there may appear new nodes with $k = 1$. Then keep on pruning these nodes until all nodes with degree $k = 1$ are removed. The k_s of the removed nodes is then set to 1. We repeat the pruning process in a similar way for the nodes with degree $k = 2$ and subsequently for higher values of k until all nodes are removed. After this process, the k-core values of all the nodes can be determined.

3.3 Measures for network robustness

We employ four structural measures to characterize the damage of a network. The *relative size of the giant connected component*, denoted as f_{GCC}, is a generally used metric to quantify the extent to which a network is damaged. Another intuitive measure is *the number of disconnected clusters* in the network. The greater the number is, the more disconnected subnetworks are due to the attack, which indicates a more serious damage. We can normalize this number by dividing it by the size of the network, denoted as $f_{cluster}$. *Network efficiency* [24] is the only topological property we adopt in this paper, which relates strongly to global shortest paths. It is defined as

$$\Lambda = \frac{1}{N(N-1)} \sum_{i,j=1, i \neq j}^{N} \frac{1}{d_{ij}}, \tag{1}$$

where N is the size of the network and d_{ij} is the length of the shortest path between nodes i and j. A lower Λ means the averaged length of shortest paths in the network is longer and the network efficiency is lower. We finally employ the *normalized susceptibility* [23], which is defined as

$$\bar{S} = \frac{\sum n_s s^2}{N}, \tag{2}$$

where n_s is the number of components of size s. A phase transition in the variation of $\bar{S}$ indicates the collapse of the network. However, the network is just shrinking if there is no phase transition during the attack.

In [13, 25], the AS-level topology of the Internet was employed as the underlying network for worm spread investigation. Hence, we also adopt two propagative measures, corresponding to the Susceptible-Infected-Susceptible (SIS) model and the Susceptible-Infected-Recovered (SIR) model, respectively, to describe the damage status of an AS-level network. For the SIS model, the nodes in the network are classified into two categories: the infected ones and the susceptible ones. Each susceptible node can be infected by its infected neighbors with a probability μ, meanwhile an infected one may return to the susceptible status with a probability β. As a result, we denote a SIS model as $SIS(\mu, \beta)$. As time evolves, the *fraction of the infected population* will eventually stabilize at a certain level, denoted as f_c^{SIS}. f_c^{SIS} can be used to characterize how far the disease can spread in the network, and thus reflect the damage status of the underlying network. All other things being equal, a smaller f_c^{SIS} implies a more severe damage. In the SIR model, a node in the network is in one of the three statuses: susceptible, infected and recovered. For a susceptible node, it may get infected by its infected neighbors with a probability μ, and an infected node may get recovered with a probability λ and will never be infected again. As a result, we denote the SIR model as $SIR(\mu, \lambda)$. Theoretically, all the nodes in the network will get recovered finally. Therefore, we utilize the *maximum fraction of nodes that get infected during the spreading process*, denoted as f_{max}^{SIR}, to characterize the worst situation.

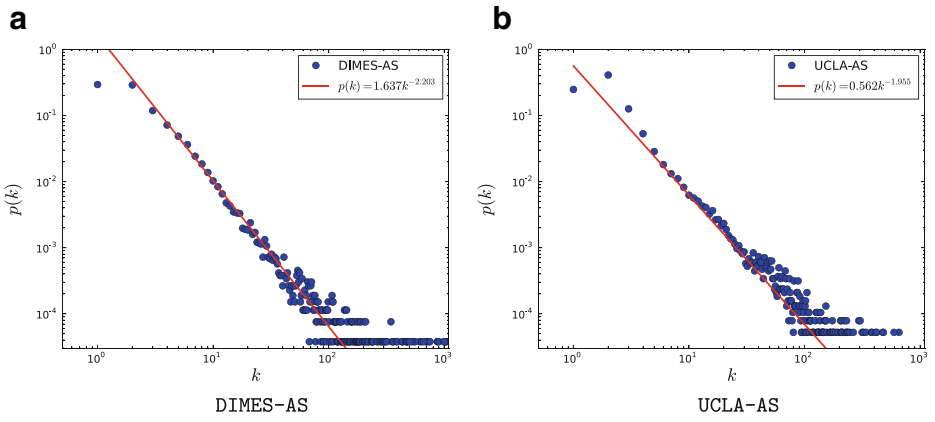

Figure 1 The degree distribution of the two data sets

In summary, we will employ four structural measures, i.e., f_{GCC}, $f_{cluster}$, Λ and $\bar{S}$, as well as two propagative measures, i.e., f_c^{SIS} and f_{max}^{SIR}, to describe the extent of the damages caused by malicious attacks.

4 Real-world network topologies

In this section, we give a brief introduction to the real-world Internet topologies used for our experiments. As a global but decentralized network [5], Internet is comprised of many smaller interconnected networks and here we investigate it mainly from the AS-level. An autonomous system (AS) is a network under the administrative control of a single organization. The routing process between different ASes is implemented by the Border Gateway Protocol, which guarantees the inter-domain routing. So if we treat an AS as a node, the path between different ASes for information exchange could be depicted as a link in the graph used to represent the Internet.

It is hard to obtain an accurate and complete picture of the AS-level Internet. In order to make our results more reliable and convincing, we use two AS-level Internet data sets. The first one, denoted as `DIMES-AS`, comes from the project of DIMES[1]. DIMES is a distributed scientific research project aiming to study the structure and topology of the Internet, with the help of a volunteer community. `DIMES-AS` was released in Mar., 2010. In this network, each node represents an AS, and each link means there exists an AS path between the related two nodes.

The second data set, denoted as `UCLA-AS`, was released by Internet Research Lab in UCLA[2] in Nov., 2010. They collected the topology from the *BGP* routing tables, the routing updates, and other existing resources. We only extract the topology from the map file released on Nov. 23, 2010.

These two are both scale-free networks, with the degree distributions shown in Figure 1. It is clear that the degree distributions follow the power-law exactly [32, 39], with $\gamma = 2.2$ for `DIMES-AS` and $\gamma = 2.0$ for `UCLA-AS`, respectively, estimated by the method given

[1] http://www.netdimes.org

[2] http://irl.cs.ucla.edu/topology/

Table 1 Real-world data sets

Data Set	$\|V\|$	$\|E\|$	$\langle k \rangle$	C	H	f_{GCC}	$f_{cluster}$	Λ
DIMES-AS	26424	90267	6.83	0.47	74.66	1.00	0.00	0.32
UCLA-AS	38200	140726	7.36	0.36	48.95	1.00	0.00	0.29

in [7]. Table 1 lists the details of the two data sets. As can be seen, while UCLA-AS contains more nodes and edges than DIMES-AS, DIMES-AS is clustered more heavily and more heterogeneous. The efficiencies of the two networks, as indicated by Λ, are similar to each other.

Nodes with the same k-core value are deemed to be in the same shell of the network. Based on the definition in Section 3, we can divide DIMES-AS into 38 shells and UCLA-AS into 77 shells. We define the fraction of nodes in each shell as f_{k_s}, where k_s is the k-core of the corresponding shell. As shown in Figure 2, f_{k_s} roughly follows a power-law distribution, which means the shell with a lower k-core contains more nodes.

5 Modeling attacks to the AS-level internet

In the section, we first give the definitions of attacks based on the degree and k-core of network nodes, respectively. Then we demonstrate how to estimate the k-core index, which enables the k-core-based attack to real-world networks.

5.1 Defining attacks

Here we focus on two kinds of malicious attacks to the AS-level Internet. One is the attack based on the node degree, called *degree-based attack* (DA). The other is the attack based on the k-core index, called *k-core-based attack* (CA). In a degree-based attack, we sort all the nodes in the descending order of degrees and remove from the network the ones with higher degrees first. Similarly, in a k-core-based attack, all the nodes in the network are ranked in the decreasing order of k-core values. Nodes located in the same shell, i.e., having a same k-core value, are further sorted in the decreasing order of degrees. Then the nodes will be removed from the highest rank to the lowest rank gradually. Note that we do not recalculate the nodes' degrees or k-core values after each wave of attack, as done in [19, 34].

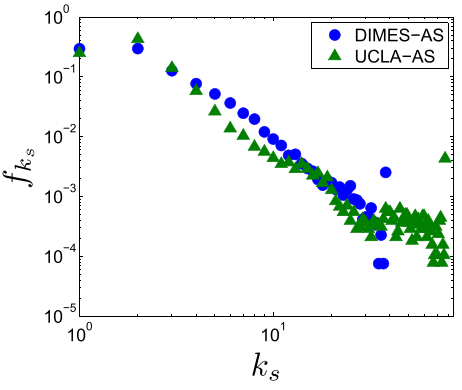

Figure 2 The fraction of the nodes in each shell

Table 2 Traceroute samples

T_{id}	1	2	3	4	5
DIMES-AS	$T(1, 500)$	$T(26, 260)$	$T(53, 530)$	$T(132, 1320)$	$T(264, 2640)$
UCLA-AS	$T(1, 500)$	$T(38, 380)$	$T(76, 760)$	$T(191, 1910)$	$T(382, 3820)$
	6	7	8	9	
	$T(528, 5280)$	$T(793, 7930)$	$T(1057, 10570)$	$T(1321, 13210)$	
	$T(764, 7640)$	$T(1146, 11460)$	$T(1528, 15280)$	$T(1910, 19100)$	

5.2 Estimating the K-core index

Generally speaking, the k-core index of a node is robust, i.e., it can be estimated from limited information of the network. To illustrate this, we perform simulations of traceroute [18] on the two AS-level topologies.

In the simulation, we randomly choose the sources and destinations from the network. Each simulation is denoted as $T(s, d)$, where s is the number of sources and d is the number of destinations. For simplification, we let $d = 10s$ (since $d = 10$ is not sufficient to setup the experiment, we let $d = 500$ when $s = 1$), and adopt the typical assumption that a route obtained by traceroute is a shortest path between the source and the destination [14]. Each sample obtained from one pair of (s, d) is denoted as $G^{T(s,d)}$. Table 2 shows the nine samples for DIMES-AS and UCLA-AS, respectively.

We first investigate the correlation between the original k-core index and the new k-core index $\left(\text{denoted as } k_s^T\right)$ estimated from traceroute samples. As shown in Figure 3, for DIMES-AS with $T(528, 5280)$ and UCLA-AS with $T(764, 7640)$, most of the nodes have their k-core values estimated correctly; that is, they are located densely near the line $k_s = k_s^T$.

We also validate the robustness of k-core index by checking the attack sequence. For each sample $G^{T(s,d)}$ from $T(s, d)$, we obtain the list of nodes in the descending order of

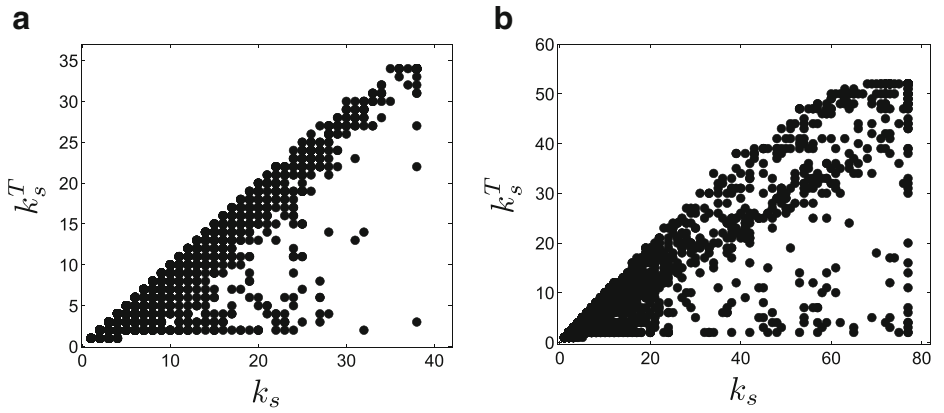

Figure 3 Correlation between k_s^T and k_s

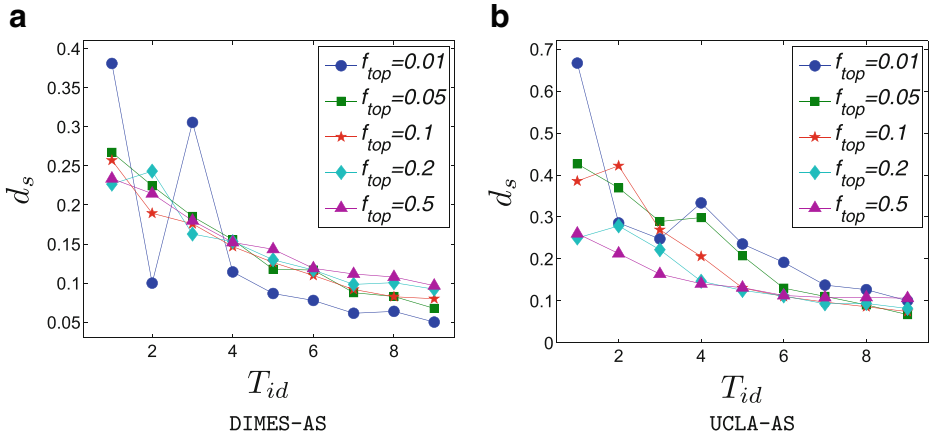

Figure 4 Distances between sample sequences and real sequences

k-core values, denoted as $\zeta^{T(s,d)}$. For the nodes with a same k-core value, we reorder them by their degrees. Similarly, from the original network we can get the attack sequence ζ. We then measure the distance between the two sequences $\zeta^{T(s,d)}$ and ζ. We define the distance between two rank lists r_1 and r_2 with a same length as follows:

$$d_s = \frac{\sum_{i \neq j} d_{ij}}{n(n-1)}, \tag{3}$$

where n is the length of the rank list, and

$$d_{ij} = \begin{cases} 1, \ r_1(i) > r_1(j), r_2(i) \leq r_2(j) \\ 1, \ r_1(i) < r_1(j), r_2(i) \geq r_2(j) \\ 1, \ r_1(i) = r_1(j), r_2(i) \neq r_2(j) \\ 0, \ \text{otherwise} \end{cases} \tag{4}$$

in which $r_1(i)$ ($r_2(i)$) stands for the rank of i in r_1 (r_2). Therefore, a lower d_s indicates the greater similarity between r_1 and r_2. We select top f_{top} nodes from $\zeta^{T(d,s)}$ as r_1, and select the same set of nodes from ζ to compose r_2. As shown in Figure 4, for both the DIMES-AS and UCLA-AS networks, as the number of sources increases, d_s decreases rapidly. For example, in DIMES-AS, the sequence from $T(528, 5280)$ is very similar to the real sequence with $d_s < 0.1$. For UCLA-AS, the sample $T(764, 7640)$ also captures most of the true sequence information.

In summary, for the real-world AS-level Internet, to estimate the k-core value of an AS or to obtain the attack sequence based on k-core index, is indeed not very difficult. For DIMES-AS or UCLA-AS, the attackers only need to collect the IP addresses of about 2 % of the total ASes to perform traceroute, which is feasible in reality.

6 Empirical study

In this section, we perform malicious attacks to real-world AS-level networks, and compare the results using the above-mentioned six measures. Some explanations will then be given to highlight the characteristics of a k-core-based attack.

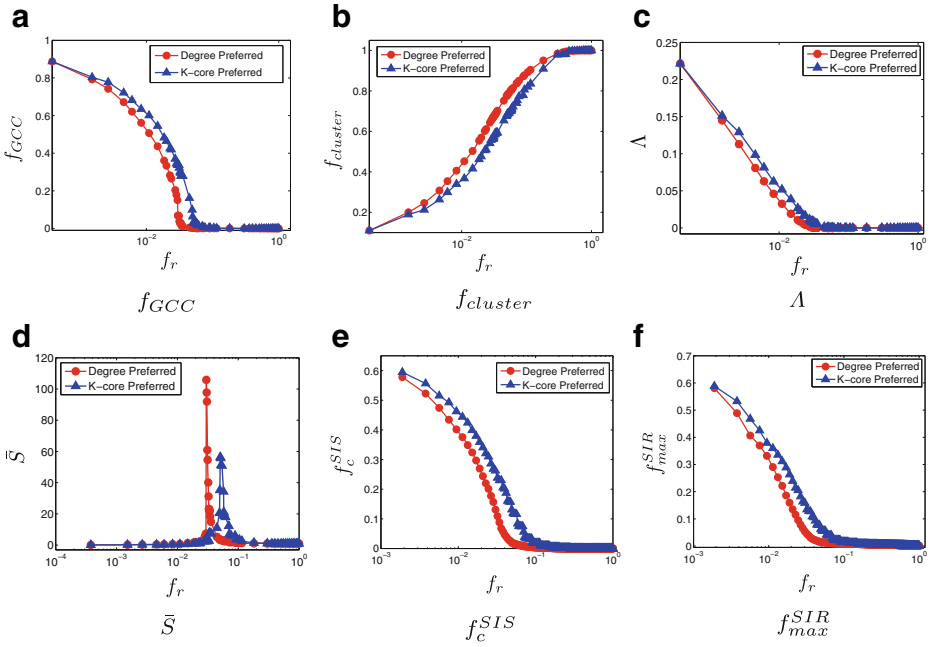

Figure 5 Comparison of two attacks to DIMES-AS

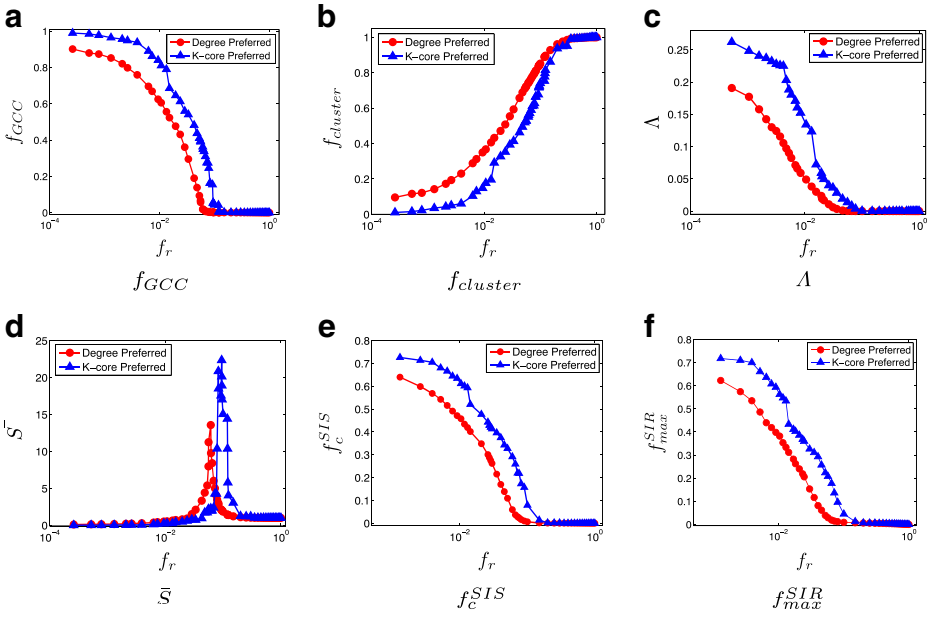

Figure 6 Comparison of two attacks to UCLA-AS

6.1 Experimental results

Here, we consider the degree-based attack and k-core-based attack. We denote the fraction of removed nodes as f_r. For the four structural measures, we first perform one round of attack and then calculate the measure values. For the two propagation measures, we first conduct one wave of attack and then simulate the SIS or SIR model on the networks for 100 times, and return the average f_c^{SIS} or f_{max}^{SIR} value. Note that we let $\mu = 1.0$ and $\beta = 0.3$ for the SIS model, and $\mu = 1.0$ and $\lambda = 0.3$ for the SIR model.

Figures 5 and 6 show the results. As can be seen, to our surprise, we find that the k-core-based attack (CA) is less malicious to the AS-level Internet than the degree-based attack (DA). We take the DIMES-AS network for illustration. As shown in Figure 5a, as f_r increases, f_{GCC} decreases more slowly for CA. This means that after removing the same amount of nodes, the network damaged by CA contains a larger GCC. Meanwhile, $f_{cluster}$ increases more quickly for DA, which implies that DA is more likely to break the network into pieces. As to Λ, it decreases less steeply for CA as f_r grows. That is to say, compared with DA, CA will not degrade the network efficiency rapidly. Finally, regarding to $\bar{S}$, the critical points of f_r at which a phase transition occurs are different for CA and DA. Specifically, the critical point for CA is 0.051, a value much larger than 0.029, the critical point for DA. This implies that DA can result in an earlier collapse of the network. Indeed, additional 528 ASes need to be attacked for CA to collapse DIMES-AS, and this number rises to 1146 for UCLA-AS.

The propagative measures also validate the less maliciousness of k-core-based attack. As can be seen in Figure 5e, for the model $SIS(1.0, 0.3)$, f_c^{SIS} decreases more slowly for CA, which means that the information or disease will spread more widely in the network bearing CA rather than DA. A similar trend can be found for f_{max}^{SIR} in Figure 5f for the $SIR(1.0, 0.3)$ model. Note that we have tried different configurations of μ, β and λ for the models and obtained similar observations invariably.

All the six measures indicate similar results for the UCLA-AS network in Figure 6; that is, the k-core-based attack is less malicious than the degree-based attack. Nevertheless, it is

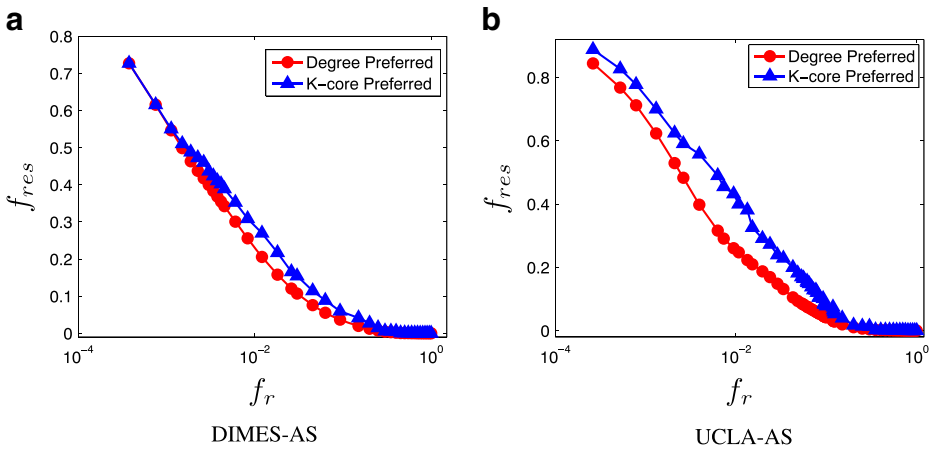

Figure 7 The fraction of residual links

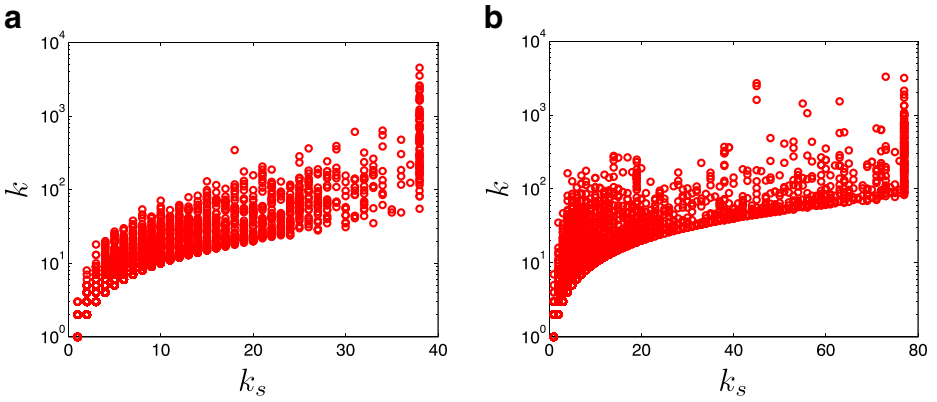

Figure 8 Comparison of degrees in the same shell

still noteworthy that the disparities of the measures between CA and DA are greater for the UCLA-AS network. For instance, as shown in Figure 6d, the critical points for CA and DA are 0.095 and 0.061, respectively, which lead to a gap larger than the one for the DIMES-AS network.

In summary, while being influential to information diffusion [22], the concept of k-core seems not very important to malicious attacks. In particular, the k-core-based attack is less malicious than the simple degree-based attack to the AS-level Internet.

6.2 Explanations and discussions

Here, we explain the above finding by exploring the interrelationship between the degree and the k-core of a node.

The essence of attacks based on node removals is to delete the links connected to those nodes. We define the fraction of residual links in the network as f_{res} and observe how it

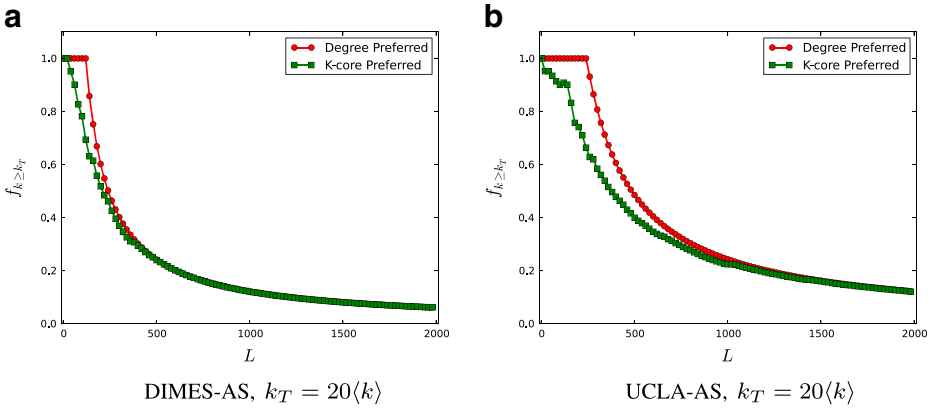

Figure 9 Comparison of $f_{k\geq k_T}$

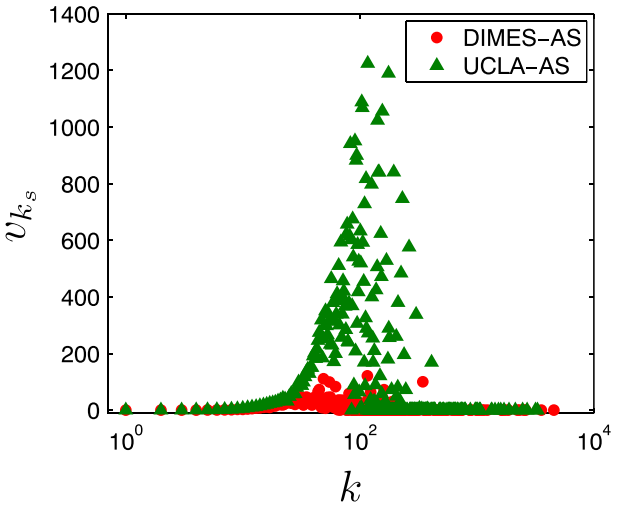

Figure 10 Variance of k-core for nodes with a same degree

varies as f_r increases. As shown in Figure 7, the fraction of residual edges for the k-core-based attack is clearly larger than the one for the degree-based attack. This implies that the k-core-based attack leads to much less link deletions.

Let us take a closer look at the nodes with a same k-core value. As shown in Figure 8, for nodes in the same shell, their degrees vary dramatically. As a result, compared with the degree-based attack, the k-core-based attack tends to delete less links from the nodes that have higher k-core values but lower degrees. To further illustrate this, we compare the attack sequences of CA and DA. We choose the first L nodes from the sequences and examine the fraction of nodes with degrees no less than a threshold k_T, denoted as $f_{k\geq k_T}$. As shown in Figure 9, compared with the degree-based attack, $f_{k\geq k_T}$ is obviously less for

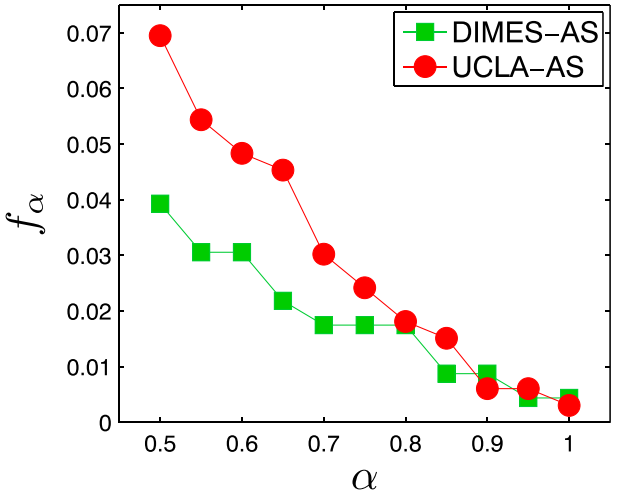

Figure 11 Variance of f_α with α

the k-core-based attack, especially at the early stage when $50 < L < 300$ for DIMES-AS or $50 < L < 1000$ for UCLA-AS. It is also noteworthy that the gap of $f_{k \geq k_T}$ is larger for UCLA-AS, which leads to the more significant differences between the two attacking strategies in Figures 5 and 6, respectively.

Moreover, to understand why the gap between the two different attacks is more obvious for UCLA-AS, we examine the variance of k-core for nodes with a same degree. Figure 10 shows the result. As can be seen, the variance in UCLA-AS is much higher than the variance in DIMES-AS, especially when $50 < k < 200$. This implies that the attack sequences by DA and CA are more inconsistent in UCLA-AS, which eventually leads to significantly different attacking effects.

Further let $v_{k_s}^{max} = \max_k v_{k_s}$, we can then generate different variance regions from the maximum value, e.g., $\left[\alpha v_{k_s}^{max}, v_{k_s}^{max}\right]$, where $0 \leq \alpha \leq 1$. For each region generated by α, we count the fraction of degrees whose variances of k-core index are in this region,

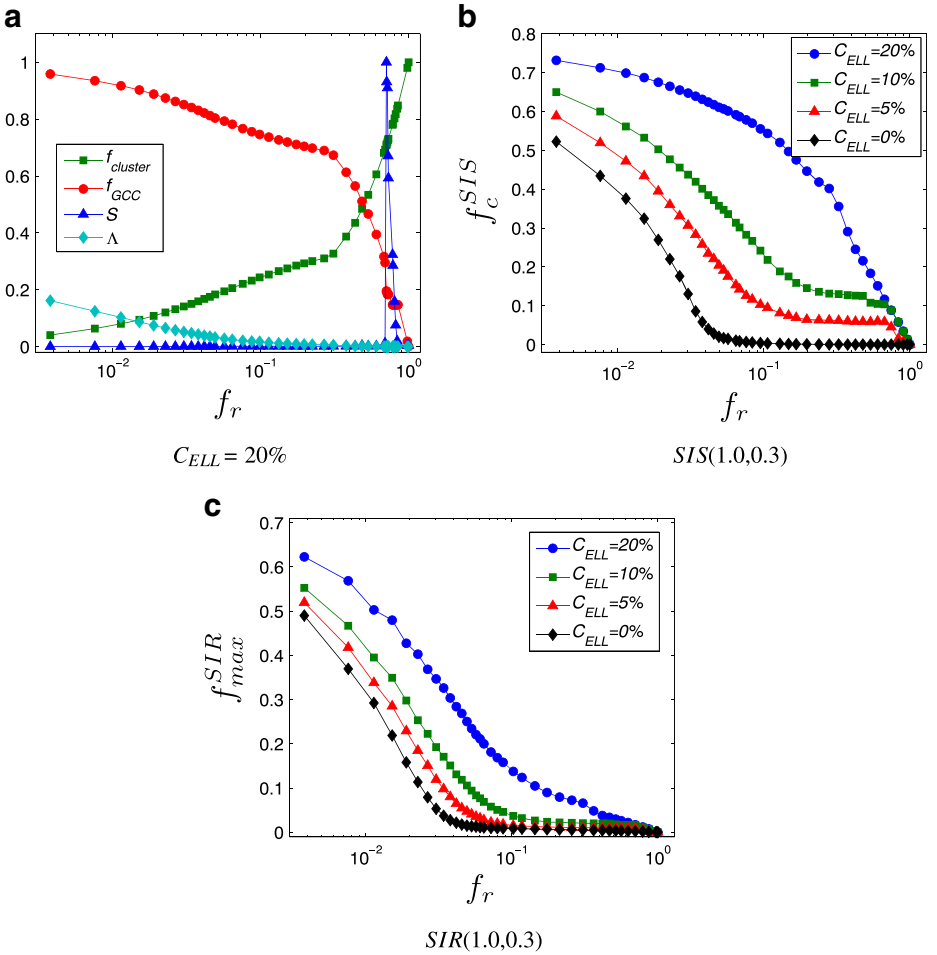

Figure 12 ELL for DIMES-AS

denoted as f_α. As can be seen from Figure 11, f_α is larger in UCLA-AS for the same α, which is consistent with our former statement that the k-core variance is more evident in UCLA-AS. It is also apparent that as α increases, the corresponding region of k shrinks, and f_α decreases accordingly. Corresponding to each variance region, we could get a window of degree at k-axis in Figure 10. For example, when $\alpha = 0.9$, the window is $[49, 116]$ for DIMES-AS, and $[117, 176]$ for UCLA-AS. When the critical value $\alpha = 1$, i.e., the maximum variance is reached, we get $k = 116$ for DIMES-AS and $k = 117$ for UCLA-AS. From the above observations, it could be learned that for the nodes with degrees between 50 and 200, the variance of k-core is tremendously large. This implies that the k-core-based attack may select some nodes with higher k-core indexes but with degrees located in [50, 200], which would eventually lead to less link deletions during the attack.

In summary, the reason for the k-core-based attack being less malicious is that the nodes with high k-core values may own low degrees, and thus leads to less link deletions in the early stage of the attack.

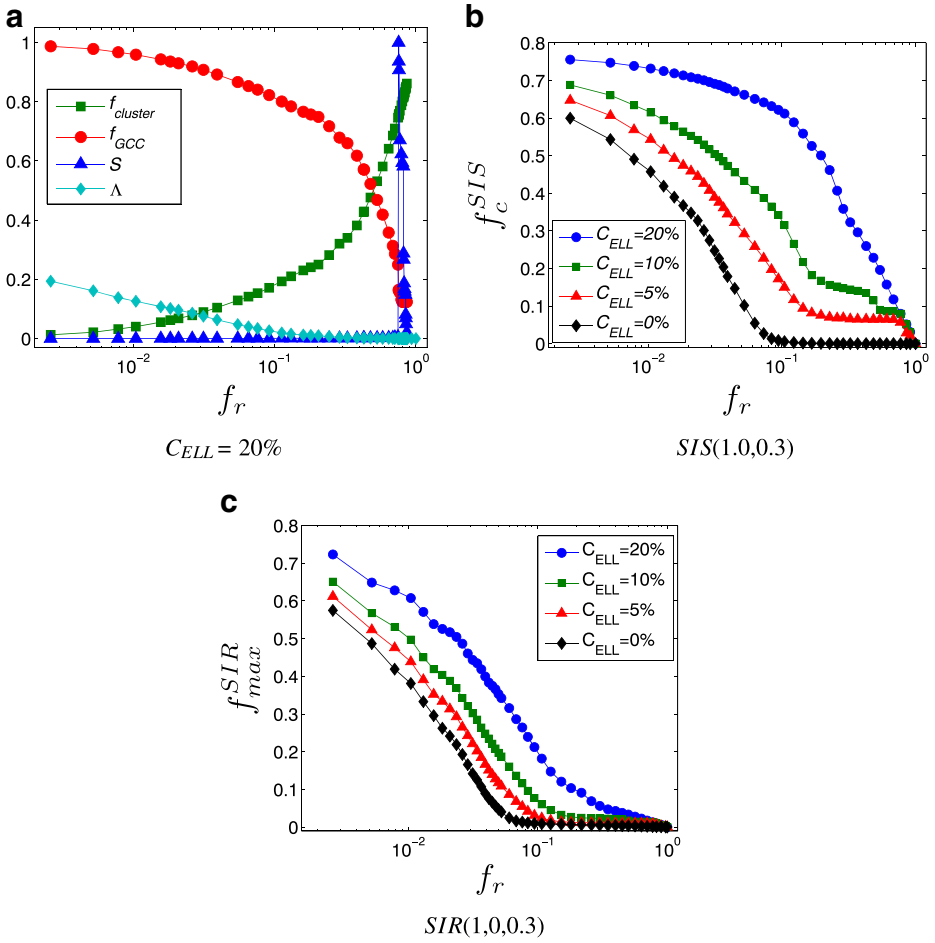

Figure 13 "ELL" for UCLA-AS

7 Enhancing the AS-level internet

The AS-level Internet is fragile to the malicious attack, although the attack may be as simple as the degree-based attack. As shown in Figures 5 and 6, we only need to remove nearly 3 % and 6 % of ASes, respectively, to make both of the two networks crashed.

Considering the extreme vulnerability of the Internet, it is intuitive to find a feasible but simple way to enhance its robustness against the malicious attack. In fact, many approaches for enhancing scale-free networks have been proposed recently, including the one called "ELL" proposed in our previous work [48], which adds new links between the nodes with lower degrees preferentially. However, the validation of "ELL" in the previous work was just performed on small-scale networks with only one robustness measure GCC. Hence, it is still interesting to test whether this strategy could work for large-scale networks in terms of more robustness metrics related to, for example, virus propagation. We define the cost of the enhancement as the fraction of newly added links as follows:

$$C_{ELL} = \frac{|E_{new}|}{|E_{origin}|}, \tag{5}$$

where E_{new} is the set of newly added links and E_{origin} is the set of links in the original network. Figures 12 and 13 show the effects of "ELL" on enhancing the two networks against the degree-based attacks. As can be seen, all the six measures indicate that the attack survivability of the AS-level Internet has been improved greatly as C_{ELL} goes up. For instance, for the enhanced `UCLA-AS` network with $C_{ELL} = 20$ %, the critical point of the phase transition has been postponed from 0.061 to 0.754, which is indeed quite an impressive improvement.

In summary, the AS-level Internet is vulnerable to simple attacks such as the degree-based attack. As a simple enhancement strategy, "ELL" can help improve the network robustness at very small costs.

8 Conclusion and future work

The Internet plays a vital role in modern communications. However, as a typical instance of scale-free networks, it is fragile to the malicious attacks. In this paper, we proposed k-core-based attack, a malicious attack preferentially to the nodes with higher k-core values, and compared it with the classic degree-based attack. Extensive experiments on two AS-level Internet topologies using six measures demonstrate that: (1) The k-core-based attack is feasible in real-world scenarios; (2) The k-core-based attack is less malicious than the degree-based attack; (3) The nodes in a same shell may have drastically varying degrees, which degrades the efficiency of a k-core-based attack; (4) As a simple scheme, "ELL" can well enhance the robustness of the AS-level Internet by connecting low-degree nodes preferentially.

In the future, we plan to extend this study in the following directions. First, we will try to give theoretical rather than empirical explanations to the less malignity of the K-core-based attack. Second, current investigation of the Internet robustness is mainly focused on the network structure, and little attention has been paid to Internet security from the view of information diffusion. It therefore would be interesting to build an attack model taking the information spread into consideration. Third, we have mentioned that natural disasters could cause damage to certain parts of the Internet and finally affect the entire network.

However, the dynamics of how failure spreading geographically from one part to the others in the Internet remains unclear. We would like to reveal it in the future research.

Acknowledgments This work was partially supported by the fund of the State Key Laboratory of Software Development Environment under Grant SKLSDE-2011ZX-02, by the Research Fund for the Doctoral Program of Higher Education of China under Grant 20111102110019, and by the National 863 Program under Grant 2012AA011005 and SS2014AA012303. Jichang Zhao thanks the China Scholarship Council (CSC) and Innovation Foundation of BUAA for PhD Graduates (YWF-13-A01-26) for support. Junjie Wu was supported in part by the National Natural Science Foundation of China under Grants 71171007, 71322104 and 70901002, by the National Information Security Research Plan of China under Grant 2012A137, by the Foundation for the Author of National Excellent Doctoral Dissertation of PR China under Grant 201189, and by the Program for New Century Excellent Talents in University under Grant NCET-11-0778.

References

1. Albert, R., Jeong, H., Barabási, A.L.: Error and attack tolerance of complex networks. Nature **406**(6794), 378–382 (2000)
2. Alvarez-Hamelin, J.I., Dall'Asta, L., Barrat, A., Vespignani, A.: K-core Decomposition: A Tool for the Visualization of Large Scale Networks. arXiv:cs/0504107v2 (2005)
3. Boguñá, M., Papadopoulos, F., Krioukov, D.: Sustaining the internet with hyperbolic mapping. Nat. Commun. **1**(62) (2010)
4. Brandes, U.: A faster algorithm for betweenness centrality. J. Math. Sociol. **25**, 163–177 (2001)
5. Butler, K., Farley, T., McDaniel, P., Rexford, J.: A survey of bgp security issues and solutions. Proc. IEEE **98**, 100–122 (2010)
6. Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y., Shir, E.: A model of internet topology using k-shell decomposition. PNAS **104**(27), 11150–11154 (2007)
7. Clauset, A., Shalizi, C.R., Newman, M.: Power-law distributions in empirical data. SIAM Rev. **51**, 661–703 (2009)
8. Clegg, R.G., Cairano-Gilfedder, C.D., Zhou, S.: A critical look at power law modelling of the internet. Comput. Commun. **33**, 259–268 (2010)
9. Cohen, R., Erez, K., Ben-Avraham, D., Havlin, S.: Resilience of the internet to random breakdowns. Phys. Rev. Lett. **85**(21), 4626–4628 (2000)
10. Cohen, R., Erez, K., Ben-Avraham, D., Havlin, S.: Breakdown of the internet under intentional attack. Phys. Rev. Lett. **86**(16), 3682–3685 (2001)
11. Cohen, R., Havlin, S., ben Avraham, D.: Efficient immunization strategies for computer networks and populations. Phys. Rev. Lett. **91**(24), 247,901 (2003)
12. Con-ed steals the internet. http://www.renesys.com/blog/2006/01/coned_steals_the_net.shtml
13. Cowie, J., Ogielski, A.T., Premore, B.J., Yuan, Y.: Internet worms and global routing instabilities. Proc. SPIE **4868** (2002)
14. Donnet, B., Friedman, T.: Internet topology discovery: a survey. IEEE Commun. Surv. Tutor. **9**(4), 2–15 (2007)
15. Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: *k*-core organization of complex networks. Phys. Rev. Lett. **96**, 040601 (2006)
16. Everett, M., Borgatti, S.P.: Ego network betweenness. Soc. Netw. **27**, 31–38 (2005)
17. Garas, A., Argyrakis, P., Rozenblat, C., Tomassini, M., Havlin, S.: Worldwide spreading of economic crisis. New J. Phys. **12**(11), 113043 (2010)
18. Guillaume, J.L., Latapy, M., Magoni, D.: Relevance of massively distributed explorations of the internet topology: qualitative results. Comput. Netw. **50**, 3197–3224 (2006)
19. Holme, P., Kim, B.J., Yoon, C.N., Han, S.K.: Attack vulnerability of complex networks. Phys. Rev. E **65**(5), 056109 (2002)
20. Huffaker, B., Plummer, D., Moore, D., Claffy, K.C.: Topology discovery by active probing. In: SAINT-W '02, pp. 90–96 (2002)
21. Kil, H., Oh, S.C., Elmacioglu, E., Nam, W., Lee, D.: Graph theoretic topological analysis of web service networks. World Wide Web **12**(3), 321–343 (2009)
22. Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A.: Identification of influential spreaders in complex networks. Nat. Phys. **6**, 888–893 (2010)

23. Kumpula, J.M., Onnela, J.P., Saramäki, J., Kaski, K., Kertész, J.: Emergence of communities in weighted networks. Phys. Rev. Lett. **99**(22), 228701 (2007)
24. Latora, V., Marchiori, M.: Efficient behavior of small-world networks. Phys. Rev. Lett. **87**(19), 198701 (2001)
25. Liljenstam, M., Yuan, Y., Premore, B.J., Nicol, D.M.: A mixed abstraction level simulation model of large-scale internet worm infestations. In: MASCOTS '02, pp. 109–116 (2002)
26. Mahadevan, P., Krioukov, D., Fomenkov, M., Dimitropoulos, X., Claffy, K.C., Vahdat, A.: The internet as-level topology: three data sources and one definitive metric. SIGCOMM Comput. Commun. Rev. **36**, 17–26 (2006)
27. Marsden, P.V.: Egocentric and sociocentric measures of network centrality. Soc. Netw. **24**, 407–422 (2002)
28. Musial, K., Budka, M., Juszczyszyn, K. Creation and growth of online social network: how do social networks evolve? World Wide Web (2013). doi:10.1007/s11280-012-0177-1
29. Musial, K., Kazienko, P.: Social networks on the internet. World Wide Web **16**, 31–72 (2013)
30. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. Phys. Rev. Lett. **86**(14), 3200–3203 (2001)
31. Pastor-Satorras, R., Vespignani, A.: Immunization of complex networks. Phys. Rev. E **65**(3), 036104 (2002)
32. Qian, T., Li, Q., Srivastava, J., Peng, Z., Yang, Y., Wang, S.: Exploiting small world property for network clustering. World Wide Web (2013). doi:10.1007/s11280-013-0209-5
33. Rekhter, Y., Li, T., Hares, S.: A border gateway protocol 4 (bgp-4). In: RFC, pp. 4271 (2006)
34. Schneider, C.M., Moreira, A.A., Andrade, J.S. Jr., Havlin, S., Herrmann, H.J.: Mitigation of malicious attacks on networks. PNAS **108**(10), 3838–3841 (2011)
35. Seidman, S.B.: Network structure and minum degree. Soc. Netw. **5**, 269–287 (1983)
36. Serrano, M.A., Krioukov, D., Boguñá, M.: Percolation in self-similar networks. Phys. Rev. Lett. **106**(048), 701 (2011)
37. Shakkottai, S., Fomenkov, M., Koga, R., Krioukov, D., Claffy, K.: Evolution of the internet as-level ecosystem. Eur. Phys. J. B **74**, 271–278 (2006)
38. Sterbenz, J.P., Hutchison, D., Cetinkaya, E.K., Jabbar, A., Rohrer, P.J., Schöller, M., Smith, P.: Resilience and survivability in communication networks: strageties, principles, and survey of disciplines. Comput. Netw. **54**, 1245–1265 (2010)
39. Sun, X., Hai, Z.: Modeling and navigation of social information networks in metric spaces. World Wide Web (2013). doi:10.1007/s11280-012-0199-8
40. Tanizawa, T., Paul, G., Cohen, R., Havlin, S., Stanley, H.E.: Optimization of network robustness to waves of targeted and random attacks. Phys. Rev. E **71**(4), 047101 (2005)
41. Wuchty, S., Almaas, E.: Peeling the yeast protein network. Proteomics **5**, 444–449 (2005)
42. Xiao, S., Xiao, G.: On imperfect node protection in complex communication networks. J. Phys. A: Math. Theor. **055**(5), 101 (2011)
43. Yan, G., Eidenbenz, S., Thulasidasan, S., Datta, P., Ramaswamy, V.: Criticality analysis of internet infrastructure. Comput. Netw. **54**, 1169–1182 (2010)
44. Youtube hijacking: A ripe ncc ris case study. http://www.ripe.net/internet-coordination/news/industry-developments/youtube-hijacking-a-ripe-ncc-ris-case-study
45. Youtube. http://www.youtube.com
46. Zhang, G.Q., Zhang, G.Q., Yang, Q.F., Cheng, S.Q., Zhou, T.: Evolution of the internet and its cores. New J. Phys. **10**(12), 123027 (2008)
47. Zhang, J., Zhao, H., Xu, J., Liu, Z.: Characterizing and modeling the internet router-level topology - the hierarchical features and hir model. Comput. Commun. **33**, 2001–2011 (2010)
48. Zhao, J., Xu, K.: Enhancing the robustness of scale-free networks. J. Phys. A: Math. Theor. **42**(19), 195003 (2009)
49. Zhou, S., Mondragón, R.J.: Accurately modeling the internet topology. Phys. Rev. E **70**, 066108 (2004)