Towards Analyzing and Improving Service Accessibility under Resource Enumeration Attack

Xiao Wang¹,b,c,d,*, Jinqiao Shi¹,a,c, Li Guo¹,a,c

¹Institute of Information Engineering, CAS, Beijing, China
²Institute of Computing Technology, CAS, Beijing, China
³National Engineering Laboratory for Information Security Technologies, China
⁴Graduate University, CAS, Beijing, China

Abstract

On the Internet, service accessibility is directly related to the reachability of entry points to its service infrastructure. There exists a main challenge in the distribution of these entry points: how to disseminate them widely while avoid the adversary from enumerating them. This paper presents the model of distribution problem under enumeration attacks. A few factors that can influence service accessibility are also presented based on a probabilistic analysis of the model. To improve accessibility, we propose dynamic distribution where the distribution parameter (number of resources assigned to each user) can be adjusted dynamically according to the confrontation situation. Experiments over simulation validate the effectiveness of this idea in accessibility improvement.

© 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of the organizers of the 2013 International Conference on Information Technology and Quantitative Management

Keywords: Accessibility, Enumeration Attack, Resource Distribution

1. Introduction

Along with its rapid development in recent years, Internet has become an important infrastructure of our daily life. As a result, cyberspace security has attracted more and more attention from both industrial and academic fields. Among all these security issues, service accessibility is one of the most fundamental and essential problem. Effective attacks such as DDoS, domain hijacking are preformed to degrade the accessibility of a targeted information system. Besides, the wide adoption of Internet censorship¹ also makes the accessibility situation even worse.

There exists many techniques and practical tools that can help users to circumvent censorship and to improve accessibility. For example, anonymous communication systems like Tor [1] and JAP [2] can help users get access to restrictive websites indirectly via a series of proxies. However, most research into improving accessibility is based on the idea of route redirection and relies on an underlying circumvention proxy infrastructure. The infrastructure provide users indirect access to restrictive services by relaying the communication between them. Unfortunately, the openness of the circumvention infrastructure leads to the following situation: on the one hand, the access points (e.g., relay IP, proxy web URL) to the infrastructure should be advised to a large number of users;

¹Tel.: +86-10-82546725; fax: +86-10-82546701.
E-mail address: wangxiao@nelmail.iie.ac.cn, shijinqiao@iie.ac.cn.
¹http://opennet.net/
on the other hand, the adversary can pretend as a normal user and discover these access points then block them. The main challenge here is to distribute access points to every user while avoid adversaries from enumerating all of them.

A bunch of resource (referred to as the access points to the circumvention proxy infrastructure) distribution strategies have been designed with the purpose of fighting enumeration attacks [3] [4] [5] [6] [7] [8] [9] [10]. For example, Tor only gives the same 3 bridges to users from a given IP subnet within a given time interval [3]. Recently, to make enumeration attacks more difficult, it requires users to solve an image CAPTCHA before requesting for bridges [11]. Other well-designed strategies such as keyspace-hopping [5] are also proposed for the purpose of bounding the enumeration capability of a single user. Besides, social relationships are employed in the design of resource distribution mechanisms [6] [7] [8] [10] to limit adversary’s ability of creating many valid Sybil identities. We noticed that, all these strategies are designed with a threat model that the attacker has bounded capability. When facing a powerful attacker with many valid user identities and large computing/human investment, these distribution strategies may not perform as well as expected.

Motivated by the possibility of adjusting distribution schemes dynamically with consideration of attack’s ability, this paper steps away from trying to design strategies that outperform current work. Instead, we investigate into a simple resource distribution strategy and study factors that may influence service accessibility. We also present dynamic resource distribution, where the distribution factor can be adjusted according to the estimate for confrontation situation. Our work provides a supplemental idea for the design of distribution strategies, and the proof-of-concept experiments validate the effectiveness of this idea. The main contribution of this paper can be summarized as follows:

- The model of resource distribution problem under enumeration attacks is presented (see Sec. 3.1). We also give the formalization and evaluation metric for resource distribution strategies based on this model (see Sec. 3.2).
- Based on the reachability of distributed resources, this paper gives a method to estimate the number of adversary’s Sybil identities (see Sec. 4.3.1). According to the estimate, a scheme to adjust the resource distribution strategy dynamically is proposed for the purpose of accessibility improvement (see Sec. 4.3.2).

2. Background and Related Work

Research into resource distribution has had a long history aiming to improve the accessibility of anonymous communication systems, covert communication systems, anti-censorship systems, etc.. The key challenge here is that: how can users discover resources efficiently without having the adversary discover and block them all.

Tor [1] is one of the most popular low-latency anonymous communication systems in the world. Tor relays form the basis of Tor network and traffic between the sender and receiver is forwarded by them in a hop-by-hop fashion. The list of all relays is updated and published once an hour, which can help Tor users to choose proper relays and get anonymous communication services. However, the adversary can also get the relay list and block access to these relays. In order to improve accessibility, Tor introduced bridge as a blocking-resistant design [3]. Tor bridges can serve as the first hop to access the core Tor network and there exists no public list of them. They are distributed in an out-of-band manner with well-designed distribution strategies. It gives only 3 bridges to a user as identified by a unique IP address subnet or email account. To enumerate all bridges, the adversary should have access to many IP address subnets or email accounts. Recently, to make enumeration attacks more difficult, it requires users to solve an image CAPTCHA before requesting for bridges [11]. Similarly, another popular anonymous communication system JAP [2] also introduced an accessibility-improvement component in its network, called forwarder [4]. It also adopts CAPTCHA in the distribution of the forwarder nodes [4].

Infranet [12] is a covert communication system that enables user to get access to restrictive sites via cooperation web servers distributed across the global Internet. A proxy discovery mechanism called keyspace hopping is introduced to distribute access points to Infranet [5]. This mechanism identifies every client and proxy with a globally unique identifier. Each client can only discover a small fraction proxies whose identifiers most closely follow a index computed from that client’s identifier. This design can efficiently limits the number of resources a Sybil identity can learn from Infranet.
Recently, some social relationship-based resource distribution strategies were proposed to limit adversary’s ability of creating many Sybil identities. Besides, the underlying social relationship can also help to manage user’s trust or even detecting potential Sybil attackers [13] [14] [15]. For example, Psiphon [8] is a web-based proxy that provides users indirect access to restrictive sites. New users can make a registration and learn the proxy address only through an invitation by an existing registered user. Another circumvention system named Kaleidoscope [7] disseminates proxy addresses only to a few high-trust people. These high-trust people perform random walks based on social relationship to advertise their node addresses which can serve as an internal relay for others. The state-of-art distribution strategies Proximax [6] and rBridge [10] also leverage social networks for proxy distribution, aiming to distribute proxies to trusted users efficiently.

To sum up, existing strategies are designed to find a trade-off between the two conflicting goals in resource distribution: disseminating resources widely and preventing adversaries from enumerating them. None of them are trying to survive a powerful adversary who has many valid identities and large computing/human investment. This forms both the basis of and motivation for this paper. We investigate factors that may influence service accessibility and propose dynamic resource distribution, which can adjust the distribution factor based on the evaluation of the confrontation situation.

3. Accessibility under Resource Enumeration Attack

3.1. Resource distribution model

Fig. 1 gives a general model of the resource distribution problem. There are two types of players in the model: a distributor who distributes resources of the circumvention proxy infrastructure (abbreviated as infrastructure in the reminder of this paper); users who requests resources from the distributor. Every user plays the role of either an honest user or an adversary’s Sybil identity.

Let $U = \{u_1, u_2, \cdots, u_n\}$ to be the user set of the infrastructure, $n$ denotes the total number of users. $U_H$ and $U_S$ are used to denote the set of honest users and that of adversary’s Sybil identities respectively, $U_H \cup U_S = U$, $U_H \cap U_S = \emptyset$. The number of honest users and adversary’s Sybils are denoted as $n_H$ and $n_S$ respectively, $n_H + n_S = n$. $R = \{r_1, r_2, \cdots, r_m\}$ is used to denote the resource set in the infrastructure, and $m$ is the total resource number at the current time. The resource distribution strategy can be denoted as a function $D : (R, u_i) \mapsto R_i$, where $R_i = \{r_{i_1}, r_{i_2}, \cdots, r_{i_k}\}$ is the resource set $u_i$ gets from the distributor. $k$ is the distribution parameter, representing for the number of resources distributed to each user.

There are two causes that can change the status of resources in the infrastructure: resource distribution and resource update. On the one hand, the adversary can block any resources he gets from the distributor. We present $\bar{R}$ to denote resources that are not reachable from the censored region, $\bar{R} \subseteq R$. $\bar{m}$ is used to represent the number of current unreachable resources, $\bar{m} = |\bar{R}|$. On the other hand, the reachability of each resource or even the number of resources may change due to resource update such as churn, IP reconfiguration, etc.

3.2. Accessibility Measurement

Accessibility is the ability to get access to the infrastructure. Actually, a honest user $u_i$’s availability of accessing the infrastructure depends on the reachability of resources he gets from the distributor. Thus, accessibility
can be defined as the probability that there exists at least one reachable resource in $R_i$, i.e., the probability that $R_i \cap (R - \tilde{R}) \neq \emptyset$. Let $P_{r_i}$ to denote the probability of $r_i \in \hat{R}$. Then, from the perspective of the honest user $u_i$, accessibility of the infrastructure can be given by (1).

$$\text{Acc}(u_i) = 1 - \prod_{j=1}^{k} P_{r_j}, \quad r_j \in R_i$$

We provide a summary of the notations used throughout this work in Tab. 1.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U = {u_1, u_2, \cdots, u_n}$</td>
<td>The user set, $</td>
</tr>
<tr>
<td>$U_H$</td>
<td>Honest users, $U_H \subseteq U$, $</td>
</tr>
<tr>
<td>$U_S$</td>
<td>Adversary’s Sybil identities, $U_S \subseteq U$, $</td>
</tr>
<tr>
<td>$R = {r_1, r_2, \cdots, r_m}$</td>
<td>The resource set, $</td>
</tr>
<tr>
<td>$\tilde{R}$</td>
<td>Resources that are not reachable, $\tilde{R} \subseteq R$, $</td>
</tr>
<tr>
<td>$\mathcal{D} : (R, u_i) \mapsto R_i$</td>
<td>Resource distribution strategy, resources assigned to $u_i$ is $R_i$, $</td>
</tr>
<tr>
<td>$P_{r_i}$</td>
<td>The probability of $r_i \in \tilde{R}$</td>
</tr>
<tr>
<td>$\text{Acc}(u_i)$</td>
<td>Infrastructure’s Accessibility from the honest user $u_i$’s view</td>
</tr>
</tbody>
</table>

### 4. Analyzing and Improving the Accessibility of Service

#### 4.1. Accessibility Analysis

Without loss of generality, we assume the distributor adopts an uniform random distribution strategy. Upon getting a resource request from the user, the distributor will select $k$ resources from $R$ randomly and disseminate them to that user. Under this assumption, resources in $R$ are selected and distributed with the same probability with each other, so the probability of them being assigned to the adversary’s Sybils is the same (as shown in (2)). For abbreviation, $P_r$ will be used to denote this probability, i.e., the probability of any resource in $R$ is unreachable.

$$P_r = P_{r_i} = P_{r_j}, \quad \forall r_i, r_j \in R$$

For the purpose of preventing enumeration attacks, we assume that the distributor distributes the same $k$ resources to a user (like it is in the Tor bridge distribution strategy [3]) in a given time interval (e.g., 24 hours). In the same time interval, the distributor will not give new resources to users even if the resources assigned to that user are all unreachable. For simplicity of analysis, we also assume that, the distributor will cache the resource update information (such as new resource join-in, resource offline, etc.) until the end of the current time interval. At the end of current time interval, the distributor will update the resource set $R$: eliminating the offline resources and unreachable resources from $R$; adding new join-in resources to $R$. Under this assumption, resource distribution process can be viewed as consisting of a series of time intervals. We can view the distribution within a time interval as a static problem, and changes only take place between two time intervals.

We consider an aggressive adversary whose Sybil identities request resources from the distributor at the very beginning of each time interval. And the adversary is eager to block them all after receiving them. This assumption is reasonable, for in a system that makes no distinction between honest users and Sybil identities, there is no reason for an adversary to keep a resource alive for longer time. For an adversary with $n_S$ Sybil identities, the number of resources he can get from the distributor who distributes $k$ random resources from $R$ of size $m$ can be approximately given by the result of the occupancy problem [16]. The approximate number of the unreachable resources after enumeration attack in a given time interval is given by (3). And the probability of any resource being blocked and becoming unreachable is given by (4).

$$\tilde{m} = m[1 - (1 - k/m)^{n_S}]$$

$$P_r = \tilde{m}/m = 1 - (1 - k/m)^{n_S}$$
Under the above assumption, the accessibility of the infrastructure from any honest user’s view is the same, $Acc$ will be used to denote this probability for abbreviation. From (1) and (4), we can conclude that:

$$Acc = 1 - P_r^k = 1 - [1 - (1 - k/m)^{n_S}]^k \quad (5)$$

4.2. Factors that influence accessibility

In (5), there are three factors that influence the accessibility of the infrastructure.

- $n_S$. Accessibility is negatively correlated with the number of adversary’s Sybil identities. Discovering these Sybil identities and treating them as a single user may help to improve the infrastructure accessibility. The thorough investigation into Sybil defense is beyond the scope of this paper (refer to [13] [14] [15] and [17] for details of Sybil defense methods).
- $m$. Accessibility is positively correlated with the number of resources in the infrastructure. The more resources, the more difficult for an adversary to enumerate a considerable fraction of them and degrade accessibility. Some circumvent tools already realized this correlation and made their efforts to enlarge their resource numbers.
- $k$. The number of resources distributed to each user ($k$) appears twice in (5), and the relationship between it and accessibility is obscure (as discussed in the following content).

In our research, one reachable resource is enough for a user to get access to the infrastructure. So, the more resources a user receives, the more possible it can get access to the infrastructure. However, a large $k$ will also make it easy for an adversary to enumerate more resources thus degrading accessibility. Considering the following two extreme cases:

1. $k = 1$. This is the most difficult case for an adversary to enumerate a large fraction of resources. However, $k = 1$ also increases the possibility of an honest user being unlucky. If the only resource that a honest user gets is unreachable, that user will have no access to the infrastructure in this entire time interval.
2. $k = m$. In this case, every user can get a full list of resources. He can get access to the infrastructure even if there is only one reachable resource in them. But this also makes it possible for the adversary to get all resources and block them.

As a result, the distributor should select a proper value for $k$ that maximizes the accessibility of the infrastructure.

4.3. Dynamic resource distribution

To maximize accessibility, $k$ should be selected according to the confrontation situation, i.e., the total resource number and adversary’s Sybil identities. Fig. 2 presents the idea of dynamic resource distribution. At the end of the current time interval, the distributor estimates the number of Sybil identities in this time interval. Then, the optimal value for $k$ is chosen based on distributor’s estimate for the confrontation situation in the following time interval.

![Fig. 2. Dynamic resource distribution. The distributor selects an optimal $k$ based on its estimate for the confrontation.](image)

4.3.1. Adversary evaluation

From the perspective of the distributor, adversary’s Sybil identities is unknown. However, (3) gives him a way to estimate it from the number of unreachable resources. The estimated Sybil identities is given by (6).

$$n_S = \log_{1-k/m}(1 - \tilde{m}/m) \quad (6)$$
At the end of the current time interval, the distributor can obtain the value of \( \hat{m} \) by probing the reachability of each resource. Besides, \( m \) and \( k \) are already known variables to the distributor. All variables in the right part of (6) are known to the distributor, so, (6) can be used to estimate the number of Sybil identities that the adversary invests in the current time interval.

### 4.3.2. Parameter optimization

As mentioned above, the distributor will update \( R \) at the beginning of a new time interval. After the update, each resource in \( R \) is online and reachable for service. And the adversary will start a new round of enumeration attack with all its Sybil identities. As a result, the new time interval can be viewed as a new confrontation situation. For the purpose of accessibility improvement, \( k \) should be chosen properly to maximize \( \text{Acc} \) as shown in (5).

The partial derivative of \( \text{Acc} \) with respect to \( k \) is given by the left part of (7). We assume the estimate of adversary’s ability in the former time interval still holds in the new confrontation situation. As a result, \( k \) is the only unknown variable in (7) to the distributor. The solution (or integers mostly near the solution) for Equation (7) is the optimal value for \( k \) that maximizes accessibility in the new confrontation situation.

\[
\frac{\partial \text{Acc}}{\partial k} = - \left[ 1 - (1 - \frac{k}{m})^{n_s} \right]^k \left\{ \ln \left[ 1 - (1 - \frac{k}{m})^{n_s} \right] + \frac{k n_s \left( 1 - k/m \right)^{n_s-1}}{m \left[ 1 - (1 - k/m)^{n_s} \right]} \right\} = 0 \tag{7}
\]

It’s not trivial to solve (7). Moreover, the cost of enumerating every possible value of \( k \) (from 1 to \( m \)) and selecting the one that maximize \( \text{Acc} \) in (5) is acceptable. Thus, we use a brute-force algorithm for the selection of parameter \( k \) in the following experiments.

### 5. Experiment and Result

In this section, we present a few scenarios and adopt simulation to explore the effectiveness of the former presented analysis.

#### 5.1. Evaluating adversary’s Sybil identities

The first experiment is used to test the adversary evaluation method presented in Sec. 4.3.1. In this experiment, the distributor has a set of \( m = 1000 \) resources and distributes \( k = 3 \) resources randomly to any user who request for them. We assume the adversary performs an enumeration attack with \( n_S \) Sybils. After the attack, the distributor makes a estimate for the number of Sybils based on (6). In this experiment, we vary \( n_S \) from 50 to 1000, and average the results over 5 simulation runs. As shown in Fig. 3, the estimates fit perfectly along the scenarios runs.

![Fig. 3. Evaluating adversary’s Sybil identities](image-url)
5.2. Static resource distribution VS. Dynamic resource distribution

Consider the following scenario: in the first time interval, the infrastructure has $m_0$ resources; at the beginning of the second time interval (e.g., 24 hours later), the number of resources increase by 20% due to resource update; and in the third time interval, it falls to 60% of that in the second interval. $m_0$ is set to be 500, 1000, 1500 and 2000 respectively in this experiment. We assume the number of adversary’s Sybils stays stable throughout the three time intervals as $n_S = 400$. The number of users is set to be $n = 5000$, which lead $n_H = 4600$ users to be honest. Any honest user who gets at least one reachable resource from the distributor will consider the infrastructure as accessible. Thus, the accessibility in this simulation experiment is calculated as the number of honest users who can get access to the infrastructure divided by $n_H$. Accessibilities are compared under two different distribution strategies:

- **Static Resource Distribution.** The distributor distribute $k_0 = 3$ resources to each user in every time interval.
- **Dynamic Resource Distribution.** The distributor sets $k_0 = 3$ in the first time interval. And in the following two time intervals, $k$ is selected dynamically using the method presented in Sec.4.3 based on its estimate for the confrontation situation (see Fig. 2).

Fig. 4 shows how dynamic resource distribution outperform static resource distribution in improving accessibility. In the first time interval, both strategies perform as well as each other since $k$ is set to be the same ($k = 3$). In the second and third time intervals, the value of $k$ is optimized by the dynamic resource distribution strategy, thus we can observe a significant difference between $Acc$ under different strategies. For example, under the setting $m_0 = 500$, the distributor with static resource distribution strategy distributes 3 resources to each user in each time interval. However, the distributor with dynamic resource distribution strategy distributes 3 resources to each user in the first time interval, and set $k = 1$ in the second and third time intervals based on its estimate for the adversary. As a result, dynamic resource distribution strategy leads to an accessibility about 3 times as high as that under static resource distribution strategy in the third time interval of setting $m_0 = 500$. This experiment validates the effectiveness of dynamic adjustment. It shows that the idea of adjusting distribution parameters dynamically should be adopted as a supplement to current distribution strategies.

![Fig. 4. Static resource distribution VS. Dynamic resource distribution. The accessibility under both static and dynamic resource distribution strategies is shown in this figure with different point shapes. The optimal value of $k$ selected in each time interval is also depicted.](image)

6. Conclusion and Future Work

Resource distribution is an essential problem in circumvention proxy infrastructures. This paper gives the model of resource distribution problem under enumeration attacks. Based on this model, our probabilistic analysis shows how a few factors in resource distribution can influence the infrastructure’s accessibility. We also present the idea of dynamic resource distribution, where the distribution parameter can be adjusted based on distributor’s
estimate for the confrontation situation dynamically. The experiments validate the effectiveness of this idea and show its potential in improving accessibility.

We will further our work to combine this idea with the state-of-art algorithms to design more practical resource distributing strategies. Thus, research can be carried out in the following two possible aspects: (i) Studying the resource distribution model of infrastructures with frequently updated resources like [18] [19]. (ii) Exploring practical methods to distinguish honest users from adversary’s Sybils [20] [13] [14] [15]. Besides the resource enumeration attack, we will consider the problem of accessibility improvement under other practical attacks in the future, e.g., resource harvesting [21] and service denying attacks[22].

Acknowledgements

This work is supported by National Natural Science Foundation of China (Grant No.61100174, 61272500), National High Technology Research and Development Program of China, 863 Program (Grant No.2011AA010701, 2011AA01A103) and National Key Technology R&D Program (Grant No.2012BAH37B04).

References

[8] Psiphon design overview 1.0 (2009).
URL http://psiphon.ca/documents/Psiphon_Design_Overview_1.0.pdf
URL https://bridges.torproject.org/