This document is published in:
IEEE Communications Letters, February 2013, 17(2), pp. 428 - 431.
DOI: 10.1109/LCOMM.2013.011113.122220

A Model to Quantify the Success of a Sybil Attack Targeting RELOAD/Chord Resources

Manuel Urueña, *Member, IEEE*, Rubén Cuevas, *Member, IEEE*, Ángel Cuevas, *Member, IEEE* and Albert Banchs *Member, IEEE*

***Abstract*—The Sybil attack is one of the most harmful security threats for distributed hash tables (DHTs). This attack is not only a theoretical one, but it has been spotted "in the wild", and even performed by researchers themselves to demonstrate its feasibility. In this letter we analyse the Sybil attack whose objective is that the targeted resource cannot be accessed by any user of a Chord DHT, by replacing all the replica nodes that store it with sybils. In particular, we propose a simple, yet complete model that provides the number of random node-IDs that an attacker would need to generate in order to succeed with certain probability. Therefore, our model enables to quantify the cost of performing a Sybil resource attack on RELOAD/Chord DHTs more accurately than previous works, and thus establishes the basis to measure the effectiveness of different solutions proposed in the literature to prevent or mitigate Sybil attacks.**

***Index Terms*—Chord, Distributed Hash Table (DHT), Kademlia, P2PSIP, REsource LOcation And Discovery (RELOAD), Sybil resource attack.**

I. Using a Sybil attack against DHT resources

A distributed hash table (DHT) is a distributed system designed to provide a simple and efficient storage and retrieval of data/resources. In this letter, we use Chord [1] as its main case of study since it is one of the most well-known DHTs, and it has been chosen as the mandatory DHT algorithm for the resource location and discovery (RELOAD) protocol [2] being standardized by the IETF P2PSIP working group. However, it must be noted that the model for the Sybil resource attack defined in this letter is also applicable to other popular DHTs such as Kademlia [3]. In Chord, the nodes participating in the DHT form a ring-shaped overlay. The position of each node in the ring is defined by its node-ID, which is the result of applying a hash function to the node's IP address and port, thus having a random nature. Similarly, each resource has an associated key that is obtained by applying the same hash function to one or more properties of the resource (typically its name). Due to resiliency reasons, each resource is replicated and stored by the r *successor* nodes of the resource's key. That is, the resource's *replicas* are in the r nodes having the subsequent IDs to the resource's key in the Chord ring.

The Sybil attack [4] consists on obtaining multiple bogus identities, called *sybils*, in order to perform different malicious actions, such as degrading the routing performance of the DHT, limiting the communications from/to one or more nodes, or blocking the access to one (or more) resources. This paper aims at developing a model that addresses the latter type of Sybil attack, targeting all the replicas of a resource. Thus, it does not consider any routing issues [5], including other attacks targeting the DHT routing, such as the Eclipse attack or Index Poisoning ones, which may also involve some kind of Sybil attack. More importantly we assume that an attacker cannot spoof or arbitrarily choose[1] its node-IDs, but that they are randomly assigned and verifiable. For instance by being provided by a trusted entity (as proposed by RELOAD [2]) or generated by some kind of verifiable cryptographic process (as in Chord [1]). Otherwise, an attacker could pretend to have the same node-ID as the targeted resource, as well as the consecutive ones, and thus the described Sybil resource attack would be quite easy to perform. Instead, we assume that the attacker must obtain multiple random node-IDs until it gets r successor IDs to the target resource's key for its sybils. Then, by controlling all its replicas, it is able to prevent the access to that resource by any peer of the RELOAD/Chord DHT.

The Sybil attack is thus a real threat to DHT systems, and not only from a theoretical point of view. Several researchers [6], [7] have reported ongoing Sybil attacks in the KAD network, and even performed it themselves for research purposes. Therefore multiple works in the literature [8] have proposed solutions to prevent or mitigate the harm produced by Sybil attacks. This letter provides a detailed model of the Sybil resource attack, and complements previous studies by providing a simple and accurate way to compute the number of node-IDs an attacker should obtain in order to perform a successful Sybil attack. Therefore, our model is a useful tool to validate the efficacy of the different solutions in practical scenarios. To the best of our knowledge, few previous studies [9]–[11] have tried to address this issue, but they just provide an approximation to the real number of attempts, which greatly overestimates its efficiency, as we will see later.

In short, the main contribution of this letter is the definition

We would like to thank the anonymous reviewers, whose comments and suggestions have help to greatly improve this letter. This work has been partially supported by the EU FP7 TREND project (257740), the Spanish T2C2 project (TIN2008-06739-C04-01) and the Madrid MEDIANET project (S-2009/TIC-1468).

M. Urueña, R. Cuevas and A. Banchs are with the Department of Telematic Engineering, Universidad Carlos III de Madrid. Av. Universidad 30, 28911 Leganés (Madrid), Spain. Email: {muruenya, rcuevas, banchs}@it.uc3m.es

A. Cuevas is with Wireless Networks and Multimedia Services Department, Institut Telecom, Telecom SudParis. 9 rue Charles Fourier, 91011 Evry, France. Email: angel.cuevas_rumin@it-sudparis.eu

A. Banchs is also with Institute IMDEA Networks. Av. del Mar Mediterraneo 22, 28918 Leganés (Madrid), Spain.

[1]Actually, this is not the case in the KAD network because clients can select their own identifiers. However, we believe that future versions of KAD clients will avoid this behavior, since it represents a serious vulnerability. For instance, version 0.49a of the eMule client has added some restrictions to limit the number of node-IDs from the same subnet.

of an analytical model that accurately specifies the number of random node-IDs that an attacker has to obtain in order to perform a Sybil resource attack with certain probability. This model is useful to quantify the vulnerability of a particular resource, as well as the whole DHT system. In particular, we use our model to evaluate the vulnerability to Sybil resource attacks of RELOAD/Chord systems, although it may be also applicable to other popular DHT systems like Kademlia.

II. Modelling the number of attempts to perform a Sybil attack against a specific resource

Let us start focusing on the simplest case, in which the target resource is stored in just one node (i.e. replica), that is, $r = 1$. In this case the goal of the attacker is to place one sybil between the target resource's key and the ID of the first successor node, which is storing the resource. We refer to the portion of the ID-space between the resource's key and its first successor's node-ID as the *attack zone*. Let us assume that the number of identifiers forming the attack zone of a given resource is z, from a total ID-space with M different identifiers. Then the attacker should generate node-IDs until it obtains one of the z identifiers within the attack zone. Since IDs have a random nature, the probability of a randomly-generated identifier belonging to the attack zone is $p = \frac{z}{M}$. Furthermore, the generation of a new ID is an independent event. Therefore, the probability of succeeding in k or less attempts is defined by the cumulative distribution function of a geometric distribution:

$$P_{resource}(z,k) = cdf_{resource}\left(\frac{z}{M}, k\right) = 1 - \left(1 - \frac{z}{M}\right)^k \quad (1)$$

In short, the previous equation defines the success probability of a Sybil attack targeting a specific resource, with an associated attack zone of size z and a single replica, after obtaining k different random IDs. However, in most DHTs the resources are typically stored in $r > 1$ nodes. In this case the attacker needs to obtain r node-IDs, instead of just one, within the attack zone. As the generation of each node-ID is an independent event, the probability of getting at least r IDs within the attack zone after k attempts is then:

$$P_{resource}(z,k,r) = \left(1 - \left(1 - \frac{z}{M}\right)^k\right)^r \quad (2)$$

Therefore, the probability of success on attacking a particular resource grows exponentially with the number of replicas (r). It is also worth mentioning that the previous expressions are also valid for Sybil resource attacks targeting other DHT systems like Kademlia/KAD, as we will see later.

III. Modelling the expected vulnerability to Sybil resource attacks of a DHT system

The previous section has described the model to calculate the probability of success of a Sybil attack targeting a particular resource with an attack zone of a specific size. This analysis may be interesting for the owner of the resource in order to check its vulnerability to Sybil attacks. However, when the resource's attack zone changes due to *churn* (i.e. nodes joining/leaving the DHT), or from the point of view of the global DHT system, the analysis for a particular attack zone is not enough. In this section we compute the probability of success of a Sybil attack targeting some random resource of the DHT. We refer to this metric as P_{sys}. This captures the expected vulnerability to Sybil resource attacks of the overall DHT system under study, and thus irrespectively of churn. Therefore, it is also a valuable metric for designing defences against Sybil resource attacks.

Toward this end, we first model the distribution of the size of the attack zones in a Chord DHT. Let us first consider the case of a single replica. In this case, if a random resource key is selected, we would like to estimate the probability that at least one successor is present in a certain attack zone with size $X \in [0, M)$, where X is a random variable representing the *attack zone size* (AZS). This probability is, by definition, equal to $1-$ probability of having no successors in X (that is, all nodes are outside X), and therefore it is $cdf_{AZS}(x,n)$. Its formal expression, where n is the number of DHT nodes, is:

$$cdf_{AZS}(x,n) = P(x < X) = 1 - \left(1 - \frac{x}{M}\right)^n \quad (3)$$

Now we can easily derive the $pmf_{AZS}(x,n)$ of the random variable X, that is, the distribution of the attack zone size in a Chord DHT with n nodes:

$$pmf_{AZS}(x,n) = \frac{\partial}{\partial x} cdf_{AZS}(x,n) = \frac{n}{M}\left(1 - \frac{x}{M}\right)^{n-1} \quad (4)$$

Note that this expression is also valid to model the distribution of the size of the attack zone for other DHTs such as Kademlia [3]. In Kademlia the nodes storing the resource are the ones with the closest ID to the resource's key (independently if such node is a successor or a predecessor of the key). Then the size of the attack zone is defined as two times the number of IDs between the resource key and the closest node-ID, so that the expression in Eq. 4 is also valid.

On the one hand, from Eq. 4 we know the probability of an attack zone of size x happening in a Chord ring with n nodes. On the other hand, Eq. 1 tell us the probability of success of a Sybil resource attack for a given attack zone size. Therefore, in order to obtain the desired metric, we simply have to perform the following integration[2] process:

$$P_{sys}(n,k) = \int_0^M P_{resource}(x,k) \cdot pmf_{AZS}(x,n)\, dx =$$

$$= \frac{n}{M}\int_0^M \left(1 - \frac{x}{M}\right)^{n-1} - \left(1 - \frac{x}{M}\right)^{k+n-1} dx = 1 - \frac{n}{n+k} \quad (5)$$

Note that Eq. 5 does not depend on the ID-space size M, but just on n, the total number of nodes in the DHT, and k, the number of node-IDs obtained by the attacker.

For the case of multiple replicas, we simply need to consider that each success process (i.e. obtaining a node-ID within the attack zone) is an independent event. Then $P_{sys}(n,k,r) = P_{sys}(n,k)^r$. This leads to the following expression for the probability of success of attacking a random resource in a DHT system where r nodes store each resource:

$$P_{sys}(n,k,r) = \left(1 - \frac{n}{n+k}\right)^r = \left(\frac{k}{k+n}\right)^r \quad (6)$$

[2]For big values of M, as is the case of most DHT systems (e.g. in RELOAD $M = 2^{128}$), the discrete ID-space $[0, M)$ can be considered a continuous one.

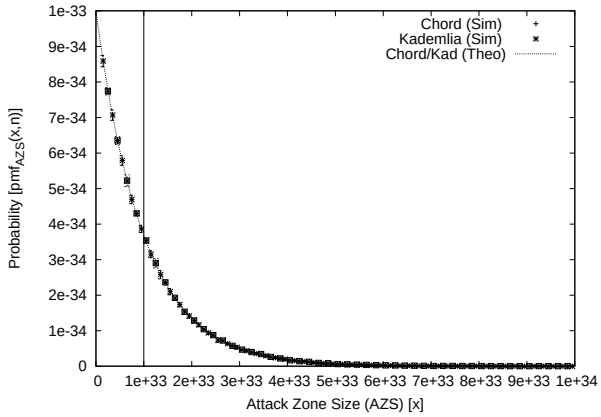

Fig. 1: Probability mass function of the attack zone size (AZS) in Chord and Kademlia (M=10^{38}, n=10^5).

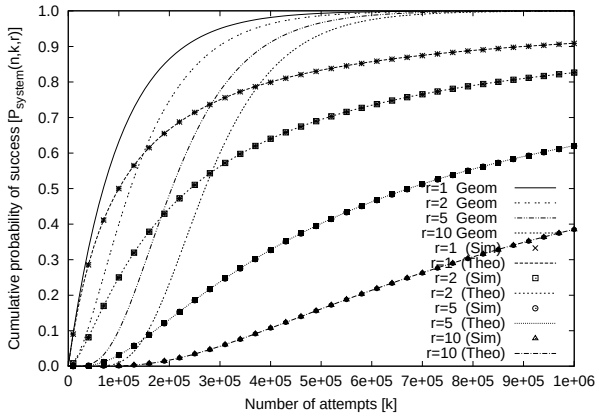

Fig. 2: Sybil resource attack attempts vs. Probability of success (M=10^{38}, n=10^5).

The above formula also allows us to know the number of sybils required to attack a DHT system with a given size and number of replicas, for a certain probability $p \equiv P_{sys}$:

$$k = p^{1/r}(k+n) = \frac{np^{1/r}}{1-p^{1/r}} = \frac{n}{p^{-1/r}-1} \qquad (7)$$

This success probability may be also seen as the portion of all resources fully controlled by an attacker with k sybils.

Finally, we would like to highlight two important issues. First, the final expressions derived above are also valid for Kademlia-based DHTs, since both $P_{resource}(z,k)$ and $pmf_{AZS}(x,n)$ are also valid for Kademlia. Second, previous works [9]–[11] have also tried to estimate the value of $P_{sys}(n,k,r)$. However, they assumed that the attack zone of all resources can be approximated by the average attack zone size, M/n. The next section discusses the consequences of this assumption.

IV. Model Validation

In this section we evaluate the accuracy of our model. In particular, we validate the distribution of the attack zone size presented in Eq. 4, and the main metric of our model, P_{sys}, defined by Eqs. 6 and 7. Note that the validation of the model for P_{sys} implicitly validates $P_{resource}$. Hence, due to space constrains we only present the results of the former.

Toward this end we have performed exhaustive simulations[3] in which we create a DHT with n=10^5 nodes randomly distributed in a M=10^{38} ID-space (emulating RELOAD's $[0, 2^{128})$ ID-space) and then reproduce Sybil resource attacks against 10^5 resources with randomly generated keys. Each simulation experiment was repeated 5 times in order to obtain the corresponding 95% confidence intervals.

In order to validate the distribution of the attack zone size (AZS) in Chord and Kademlia DHTs, for each resource we first compute the size of its attack zone, both as the distance from the resource's key to its successor node-ID (Chord), and as two times the distance to the closest node-ID (Kademlia). Fig. 1 presents the distribution of the size of the attack zones from the analytical model and the simulation experiments. We observe that the distribution of the attack zone size in Chord and Kademlia is exactly the same, and that our model accurately matches the simulation results.

Fig. 1 also depicts the average size (M/n=10^{33}) of the attack zones. Previous works [9]–[11] have assumed that the size of the attack zone for all resources is equal to this value. However, our model and simulations demonstrate that attack zones smaller than the average are likely to occur. Hence, since smaller zones are much harder to attack, our results suggest that previous works are overestimating the probability of success of the Sybil attack, because they only consider the average attack zone size, instead of the real AZS distribution as ours.

For the validation of $P_{sys}(n,k,r)$ we use the same setup (M=10^{38}, n=10^5), but now we simulate one Sybil attack against each of the R=10^5 resources in the DHT. For each resource we generate as many random IDs as needed (up to a maximum of $k_{max} = 10^6$ attempts) until succeeding in the attack, which means obtaining r (the number of replicas) IDs within the attack zone. For each Sybil resource attack we compute the number of attempts needed to succeed using different values of $r \in [1, 10]$.

Fig. 2 shows the analytical and the simulation-based cumulative distributions of $P_{sys}(n,k,r)$ with $r = 1, 2, 5$ and 10. The results demonstrate that our model is extremely accurate since simulation values overlap the curves of the model. Furthermore, Fig. 2 also presents the distribution in case of assuming that all attack zones are equal to the average attack zone size (as assumed by previous works [9]–[11]) for the same cases of $r = 1, 2, 5$ and 10, which are labeled as "*Geom*" in the figure. This confirms our previous hypothesis, since this approximation clearly leads to overestimate the probability of success of Sybil attacks, and the error becomes even bigger with more replicas. This effect can be better seen in Figure 3, which shows how the number of replicas affects the number of sybils needed to succeed with a given probability (p=0.1) for both our model (Eq. 7) and previous works based on a geometric distribution, that is, $p = (1-(1-1/n)^k)^r \Rightarrow k = log(1-p^{1/r})/log(1-1/n)$.

[3]The source code of the developed simulator can be downloaded from http://www.it.uc3m.es/muruenya/SybilResourceAttackSimulator.tgz

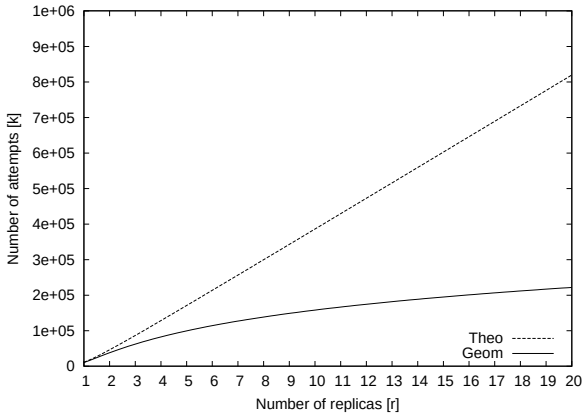

Fig. 3: Sybil resource attack attempts vs. Number of replicas for a target success probability of $p = 0.1$ ($n = 10^5$).

The suitability of replication as a defense mechanism against the Sybil resource attack may have been underestimated by previous works, since each new replica increases the number of attack attempts linearly, much more than previously thought. Moreover, the approximation error made by previous works may lead DHT system designers to overreact, and devise additional defenses against the (overestimated) Sybil attack that are too demanding for innocent users, as explained in next section.

V. Vulnerability of DHT systems to Sybil resource attacks

The goal of this section is to briefly discuss the vulnerability level of existing DHT systems to the Sybil resource attack. Some of the most popular DHT systems like KAD, the one employed by the eMule file-sharing network, or the two main DHTs associated to BitTorrent, account for millions of concurrently connected users. In particular, Steiner et al. [12], and more recently Cholez et al. [7], reported that the number of users concurrently connected to the KAD network varies between 3.3 and 4.5 millions. Furthermore, the minimum number of replicas of KAD resources is $r = 10$.

These values, together with the Eqs. 6 and 7, allow us to make an accurate estimation of the real vulnerability of DHT systems. In order to perform a Sybil resource attack on a RELOAD/Chord DHT with a similar size to KAD ($n = 4 \cdot 10^6$ and $r = 10$), an attacker would need to generate 55.7 million IDs to guarantee a success probability of 50%. The way to generate or gather such amount of node-IDs greatly varies between different DHT systems and proposals [8], ranging from changing the local port number, to leaving and joining the network, generating a new public/private key pair, or solving crypto-puzzles. Thus the number of resources or the required time to perform a Sybil attack depends on the specific mechanism to obtain node-IDs.

For instance, we can start considering a random ID generation/assignment process as simple as the one originally proposed for Chord [1], in which the node-ID is the result of a hash operation on the node IP address and port. In this case, given that each IP address can be associated with up to 65536 different ports, an attacker should have access to 851 different IP addresses. Other solutions proposed in the literature [8] to mitigate Sybil attacks, other than a centralized authentication server like in RELOAD [2], include using crypto-puzzles [11], or charging a small fee to each new user [13]. The appropriate cost of these mechanisms for legitimate users could be tuned for that particular DHT system by means of our model. For instance, by charging 10 cents (instead of 20 dollars, as proposed in [13]) per random ID to new KAD users, an attacker willing to spend 1 million dollars would only have a 3.5% probability to completely block a resource replicated in 10 nodes. On the other hand, if each node-ID requires generating a new RSA key pair or solving a crypto-puzzle that takes one second on average, 90% of all DHT resources can be targeted after almost 12 years of computing power, or just half a year with a cluster of 24 machines.

In short our results suggest that performing a Sybil resource attack on RELOAD/Chord systems, or existing DHT systems such as the KAD network (assuming secure routing and random IDs) is more difficult than previously reported, but still doable for resourceful attackers, using for instance a botnet or a cluster. Finally it is worth noting that our model can be used in order to quantify the level of security added by the different solutions proposed against Sybil attacks and compare their performance in an objective manner.

In a future work we will analyze partial Sybil resource attacks where the attacker only controls a subset of all replicas.

References

[1] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. "Chord: A scalable peer-to-peer lookup service for Internet applications". *ACM SIGCOMM'01*, San Diego (USA), Aug. 27-31, 2001.

[2] C. Jennings, B. Lowekamp, E. Rescorla, S. Baset and H. Schulzrinne. "REsource LOcation And Discovery (RELOAD) Base Protocol <draft-ietf-p2psip-base-22>". *Internet Draft*, Jul. 2012.

[3] P. Maymounkov and D. Mazires. "Kademlia: A Peer-to-Peer Information System Based on the XOR Metric". *1st Int. Workshop on Peer-to-Peer Syst. (IPTPS'02)*, Mar. 7-8, 2002.

[4] J. R. Douceur. "The Sybil attack". *1st Int. Workshop on Peer-to-Peer Syst. (IPTPS'02)*, Mar. 7-8, 2002.

[5] H. J. Kang, E. Chan-Tin, N. J. Hopper, Y. Kim. "Why Kad Lookup Fails". *9th Int. Conf. on Peer-to-Peer Computing (P2P'09)*, Sept. 9-11, 2009.

[6] M. Steiner, T. En-Najjary and E. W. Biersack. Exploiting KAD: possible uses and misuses. *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 5, pp. 65-70. October 2007.

[7] T. Cholez, I. Chrisment, O. Festor, and G. Doyen. "Detection and mitigation of localized attacks in a widely deployed P2P network". *Peer-to-Peer Netw. Appl.*. 2012.

[8] G. Urdaneta, G. Pierre, and M. Van Steen. "A Survey of DHT Security Techniques". *ACM Comput. Surveys*, vol. 43, no. 2, Feb. 2011.

[9] M. Srivatsa and L. Liu. "Vulnerabilities and Security Threats in Structured Overlay Networks: A Quantitative Analysis". *20th Annu. Comput. Security Appl. Conf. (ACSAC'04)*, Dec. 6-10, 2004.

[10] D. Cerri, A. Ghioni, S. Paraboschi and S. Tiraboschi. "ID mapping attacks in P2P networks". *IEEE Global Telecommun. Conf. (GLOBECOM'05)*, Nov. 28 - Dec. 2, 2005.

[11] R. Zhang, J. Zhang, Y. Chen, N. Qin, B. Liu and Y. Zhang. "Making eclipse attacks computationally infeasible in large-scale DHTs", *IEEE 30th Int. Performance Comput. and Commun. Conf. (IPCCC'11)*, Nov. 17-19, 2011.

[12] M. Steiner, T. En-Najjary and E. W. Biersack. "A Global View of KAD". *7th ACM SIGCOMM Conf. on Internet Meas. (IMC'07)*, Oct. 24-26, 2007.

[13] M. Castro, P. Druschel, A. Ganesh, A. Rowstron and D. S. Wallach. "Secure routing for structured peer-to-peer overlay networks". ACM SIGOPS Operating Systems Review, vol. 36, pp. 299-314, 2002.