

◎网络、通信、安全◎

# 利用 Sybil 攻击提升 PageRank 值

张大陆, 陈 晴, 胡治国

ZHANG Da-lu, CHEN Qing, HU Zhi-guo

同济大学 电信学院 计算机科学与技术系, 上海 201804

Department of Computer Science, Tongji University, Shanghai 201804, China

E-mail: Daluz@acm.org

ZHANG Da-lu, CHEN Qing, HU Zhi-guo. Using Sybil attack to optimize PageRank. Computer Engineering and Applications, 2008, 44(19): 96-97.

**Abstract:** PageRank, as the most important reputation ranking factor of Google, is prone to Sybil attack as recently research results prove. A higher ranking may offer an economic benefit. Based on some preventive efforts on finding the best Sybil strategy to manipulate PageRank on its simplest version, the paper focuses on how to use strategy to optimize PageRank into its aimed ranking. Pay attention that the optimization work is done simulative without consideration for complex network environment, so it would still be far away to put such strategy into actual implementation, which is also the future work.

**Key words:** Sybil; PageRank; reputation system

**摘 要:** 最近的研究证明 Google 的页面等级策略 PageRank 容易受到 Sybil 攻击。着眼于对如何构造 Sybil 攻击来优化 PageRank 值。提出了构造 Sybil 攻击的策略模型, 考虑节点生成成本和迭代收敛性的前提下, 证明了攻击的有效性。最后在模拟环境下对攻击进行了量化分析。

**关键词:** 女巫攻击; 页级; 信任系统

**DOI:** 10.3778/j.issn.1002-8331.2008.19.028 **文章编号:** 1002-8331(2008)19-0096-02 **文献标识码:** A **中图分类号:** TP393.08

## 1 引言

虚拟环境阻碍了用户直接了解远程实体的物理特性。作为一种替代方法, 用户只能通过表示实体信息抽象的身份来了解它。在网络中, 一个恶意节点通过模拟其他节点或只是简单使用虚假身份来使自己对外呈现为多个节点被称为 Sybil 攻击。

Sybil 攻击在 Douceur<sup>[1]</sup>中首次被提到。E.Sit 和 R.Morris 也在他们的点对点分布式哈希表的安全性考量一文中指出, 一个安全的身份配方法应使得一个节点无法任意选择它的节点标识即身份。Castro 等人<sup>[2]</sup>也认为对一个 P2P 网络来说, 防御 Sybil 攻击最根本的是节点身份的统一分发不能被攻击者控制。他们提出基于可信的权威中心进行 ID 分配的解决方案。同时, Srivatsa 和 Liu<sup>[3]</sup>也在基于外部 ID 的正式模型中定义了一个 ID Mapping 方案可以部分防御 Sybil 攻击。

当大部分的研究者将他们的对 Sybil 的分析研究工作都集中在身份分配方案上时, A.Cheng<sup>[4,5]</sup>提出对称的信任系统更容易被 Sybil 攻击。他提出许多等级系统容易被用户操控, 并且用户经常有欺骗的动机, 即在系统中获得更高信用等级。因为一个更高的等级意味着更高的经济利益。在对 ebay 信任系统的研究中也发现, 购买者宁愿去多支付 8% 的额外费给高信用的卖家。

Google 搜索引擎中使用 PageRank 进行等级评定, 它能够

帮助用户更快的了解万维网的异质性。PageRank 基于一个网络图来计算每个网页的等级, 这种方法也可以应用在 P2P 网络中。虽然 PageRank 已经被证明是一个十分有效的等级评定系统, 但它也易于被各种攻击策略如共谋或 Sybil 的操控<sup>[6]</sup>。

## 2 定义

给定一个集合  $V$ , 用有向图  $G=(V, E)$  其中边界集合  $E$  代表用户之间的直接信任关系。例如, 在网络中, 一个边界  $(v, u)$  代表从节点  $v$  出发到节点  $u$  的向外链接, 那么假定  $F_u$  是  $u$  所指向的集合,  $B_u$  是指向  $u$  的集合,  $u$  的链接数  $N_u=|F_u|$ ,  $c$  是一个正常化因子。

### 2.1 PageRank 算法

Lawrence Page 和 Sergey Brin 提出了用户行为的随机冲浪模型, 来解释上述算法。他们把用户点击链接的行为, 视为一种不关心内容的随机行为。而用户点击页面内的链接的概率, 完全由页面上链接数量的多少决定的。一个页面通过随机冲浪到达的概率就是链入它的别的页面上的链接的被点击概率的和。  $c$  可以视为用户无限点击下去的概率,  $(1-c)$  则就是页面本身所具有的网页级别。

在网络图  $G$  中 PageRank 的值由在图中随机步进的概率决定。在图中用  $c$  表示从节点  $u$  沿  $u$  的向外链接边界集以概率

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.90204010)。

作者简介: 张大陆(1949-), 男, 博士生导师, 主要研究领域为网络安全、语义网; 陈晴(1982-), 女, 硕士, 主要研究领域为网络安全。

收稿日期: 2007-12-24

修回日期: 2008-02-29

$1/N_u$  选择边界步进。这样可计算得出 PageRank 的值:

$$R(u) = (1-c) + c \sum_{v \in B_u} \frac{R(v)}{N_v} \quad (1)$$

### 2.2 策略模型

正如图 1 和图 2 所示, 黑色节点表示正是提升 PageRank 的目标节点, 白色节点是 Sybil 节点。目标页  $p$  个出链接(向外边界)和  $q$  个入链接(向内边界)。假定目标节点的入链接的 PageRank 分别是  $\rho_1, \rho_2, \rho_3, \dots, \rho_q$ 。可以证明, 当一个 Sybil 节点插入中时, 节点的数量, 链接中出链接及入链接是影响提升 PageRank 值的关键因素。

因为 Sybil 攻击的系统稳定性和不可侦测性, 现确保在攻击时使目标节点的出链接和入链接数目保持稳定, 使得在其他节点不能感知到目标节点的任何变化的同时, 优化 PageRank 值。因此, 评估策略必须考虑到生成 Sybil 节点的代价和迭代计算中 PageRank 的收敛性。

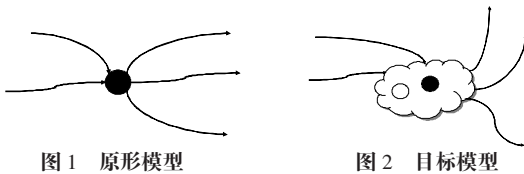


图 1 原形模型

图 2 目标模型

### 3 Sybil 策略内部构造

已经在前期的研究中证明, Sybil 策略的内部构造和最终优化结果有关。根据本文的研究结果, 如果 Sybil 的数量是  $m$ , 那么目标节点所能提升的 PageRank 的值为  $PR_{(S_i)}^j, 1 \leq i \leq m, 1 \leq j \leq n$ , 请注意  $i$  在第  $j$  次的迭代。

$$PR_{(S_i)}^j = (1-c) + c(PR_{(T)}^j / (m+p))$$

$$PR_{(T)}^j = (1-c) + c((\rho_1 + \rho_2 + \dots + \rho_q) + PR_{(S_1)}^j + PR_{(S_2)}^j + \dots + PR_{(S_m)}^j) =$$

$$(1-c) + c(\sum_{k=1}^q \rho_k + \sum_{i=1}^m PR_{(S_i)}^j)$$

第 1 次迭代:

$$PR_{(S_i)}^1 = (1-c); PR_{(T)}^1 = (1-c) + c(\sum_{k=1}^q \rho_k)$$

第 2 次迭代:

$$PR_{(S_i)}^2 = (1-c) + c(((1-c) + c(\sum_{k=1}^q \rho_k)) / (m+p)) = (1-c) + c(1 -$$

$$c) / (m+p) + c^2(\sum_{k=1}^q \rho_k) / (m+p)$$

$$PR_{(T)}^2 = (1-c) + c(\sum_{k=1}^q \rho_k + m \cdot ((1-c) + c(1-c) / (m+p)) + c^2(\sum_{k=1}^q \rho_k) /$$

$$(m+p))) = (1-c) + c \sum_{k=1}^q \rho_k + c(1-c)m + c^2(1-c)m / (m+p) +$$

$$c^3(\sum_{k=1}^q \rho_k)m / (m+p)$$

第 3 次迭代:

$$PR_{(S_i)}^3 = (1-c) + c(((1-c) + c(1-c)m + c^2(1-c)m / (m+p)) +$$

$$c \sum_{k=1}^q \rho_k + c^3(\sum_{k=1}^q \rho_k)m / (m+p)) / (m+p) = (1-c) + c / (m+p) (((1-c)$$

$$+ c(1-c)m + c^2(1-c)m / (m+p) + c \sum_{k=1}^q \rho_k + c^3(\sum_{k=1}^q \rho_k)m / (m+p))) = (1-c) +$$

$$c(1-c) / (m+p) + c^2(1-c)m / (m+p) + c^3(1-c)m / (m+p)^2 + c^2 \sum_{k=1}^q \rho_k /$$

$$(m+p) + c^4(\sum_{k=1}^q \rho_k)m / (m+p)^2$$

$$PR_{(T)}^3 = (1-c) + c(\sum_{k=1}^q \rho_k + m \cdot ((1-c) + c(1-c) / (m+p) + c^2(1-c)m /$$

$$(m+p) + c^3(1-c)m / (m+p)^2 + c^2 \sum_{k=1}^q \rho_k / (m+p) + c^4(\sum_{k=1}^q \rho_k)m / (m+p)^2)) =$$

$$(1-c) + c(1-c)m + c^2(1-c)m / (m+p) + c^3(1-c)m^2 / (m+p) + c^4(1-c)m^2 /$$

$$(m+p)^2 + c \sum_{k=1}^q \rho_k + c^3 \sum_{k=1}^q \rho_k m / (m+p) + c^5(\sum_{k=1}^q \rho_k)m^2 / (m+p)^2$$

第 4 次迭代:

$$PR_{(T)}^4 = (1-c) + c \sum_{k=1}^q \rho_k + c(1-c)m + c^2m \cdot ((1-c) + c \sum_{k=1}^q \rho_k + c(1-c)m +$$

$$c^2(1-c)m / (m+p) + c^3(1-c)m^2 / (m+p) + c^3 \sum_{k=1}^q \rho_k m / (m+p) + c^4(1-c)m^2 /$$

$$(m+p)^2 + c^5(\sum_{k=1}^q \rho_k)m^2 / (m+p)^2) / (m+p) = (1-c) + c(1-c)m + c^2(1-c)m /$$

$$(m+p) + c^3(1-c)m^2 / (m+p) + c^4(1-c)m^2 / (m+p)^2 + c^5(1-c)m^3 / (m+p)^2 +$$

$$c^6(1-c)m^3 / (m+p)^3 + c \sum_{k=1}^q \rho_k + c^3 \sum_{k=1}^q \rho_k m + c^5 \sum_{k=1}^q \rho_k m^2 / (m+p)^2 + c^7(\sum_{k=1}^q \rho_k)m^3 /$$

$$(m+p)^3; \dots$$

$$PR_{(T)}^n = (1-c) \left( \sum_{i=0}^{n-1} c^2 m^i / (m+p)^i + \sum_{i=1}^{n-1} c^{2i+1} m^{i+1} / (m+p)^i \right) +$$

$$\sum_{k=1}^q \rho_k \left( \sum_{i=1}^n c^{2i-1} m^i / (m+p)^i \right)$$

显然, PageRank 通过 Sybil 策略可以成功被提升。在另一篇文章中提出其他方案来改进该策略。

### 4 模拟和结果评估

本章通过模拟实验来评估本文的 Sybil 策略。模拟模型有 20 个入链接, 每一个节点的默认 PageRank 被设定为 1。比较在使用 Sybil 策略前后的 PageRank。策略模型如图 3 所示, Sybil 节点的出链接和入链接同时作用于目标节点。计算 PageRank 的迭代数设为 100, 如图 4 所示, 能够发现目标节点 PageRank 值被成功提高了近 2 倍。

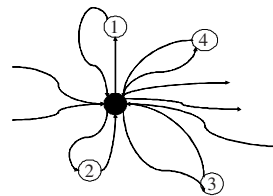


图 3 具有 4 个 Sybil 节点的模型

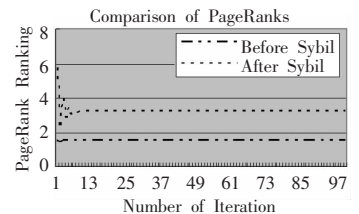


图 4 模拟实验结果

### 5 结论

本文的研究是基于 PageRank 的现行算法, 利用 Sybil 攻击来优化 PageRank 值。必须指出的是, 本文优化策略已经在模拟环境中被证实是有效的, 但本文研究成果还不能直接应用于真实的网络环境中。后期的研究工作包括在实际网络中实现