doi:10.11959/j.issn.1000-436x.2015036

Tor 匿名通信网络节点家族的测量与分析

王啸[1,2,3,4,5]，方滨兴[1,2,4]，刘培朋[1,2,3,4]，郭莉[2,4]，时金桥[2,4]

（1. 中国科学院 计算技术研究所，北京 100190；2. 中国科学院 信息工程研究所，北京 100093；
3. 中国科学院 研究生院，北京 100190；4. 信息内容安全国家工程实验室，北京 100093；
5. 国家计算机网络应急技术处理协调中心，北京 100029）

摘 要：重点关注 Tor 匿名通信系统的家族（family）设计，从连续两年的 Tor 网络真实数据中提取了数千个 Tor 节点家族，揭示了 Tor 节点家族的规模、带宽、地理分布等规律，同时也研究了超级家族背后的运营者身份。基于测量结果的分析验证了 Tor 的家族设计在保障其匿名性方面所发挥的不可替代的重要作用。相应的安全性分析说明了恶意 Tor 节点家族对 Tor 网络可用性带来的挑战，也揭露了 Tor 网络中隐藏家族现象的普遍性及其对 Tor 网络匿名性所造成的威胁。

关键词：Tor；匿名通信；节点家族；收割攻击；测量；隐藏家族

中图分类号：TP393.08 **文献标识码**：A

Measuring and analyzing node families in the Tor anonymous communication network

WANG Xiao[1,2,3,4,5], FANG Bin-xing[1,2,4], LIU Pei-peng[1,2,3,4], GUO Li[2,4], SHI Jin-qiao[2,4]

(1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;
2. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China;
3. Graduated University of Chinese Academy of Sciences, Beijing 100190, China;
4. Chinese National Engineering Laboratory for Information Security Technologies, Beijing 100093, China;
5. National Computer Network Emergency Response Technical Team/Coordination Center, 100029, China)

Abstract: Tor family was focused on design and given a family-level measurement of it. Based on Tor node families (over 3000) discovered from live Tor network data (from Jan. 2011 to Dec. 2012), a few characteristics of Tor node families were revealed, such as family size, bandwidth, geographical distribution as well as operators providing a few big families. The analysis validated the irreplaceable role played by family design in enhancing Tor's anonymity. Based on the measurement, security analysis showed the serious availability threat a compromised node family can cause to the Tor network. Besides, It also discussed Tor hidden families and the potential anonymity risk caused by them.

Key words: Tor; anonymous communication; node family; harvesting attack; measurement; hidden family

1 引言

作为目前世界上最流行的匿名通信系统，Tor[1]已经吸引了来自军队、记者、政府、活动家等各行各业的众多用户。目前，世界上拥有着超过 3 000 个 Tor 中继节点，提供了超过 1.5 Gbit/s 的转发服务带宽[2]。在 Tor 网络中，Tor 用户的通信流量由这些中继节点负责转发，保障了用户访问网络服务的匿名性。在提供匿名性的同时，Tor 网络中继节点广泛的地理分布也使之成为了一个逃避网络监管的

收稿日期：2013-10-24；**修回日期**：2014-03-25

基金项目：国家高技术研究发展计划（“863”计划）基金资助项目（2011AA010701, 2012AA013101）；国家自然科学基金资助项目（61100174）；国家科技支撑计划基金资助项目（2012BAH37B04, 2012BAH42B02）

Foundation Items: The National High Technology Research and Development Program of China (863 Program) (2011AA010701, 2012AA013101); The National Natural Science Foundation of China (61100174); The National Key Technology R&D Program (2012BAH37B04, 2012BAH42B02)

重要工具[3, 4]。

Tor 网络的基本组成如图 1 所示。Tor 目录服务器是其网络的核心，负责收集 Tor 网络中的中继节点信息并以节点快照（consensus）及节点描述（descriptor）的形式发布给 Tor 代理[5]； Tor 中继节点是 Tor 网络的基础，在 Tor 网络中的匿名通信流量都是通过由多个 Tor 中继节点所组成的匿名通信链路来转发的；Tor 代理运行于 Tor 用户端，它负责建立匿名链路并在用户的网络应用程序与 Tor 匿名链路之间中转网络流量。在图 1 中，由 3 个 Tor 中继节点构成了一条 Tor 匿名通信链路，这 3 个节点依据其位置依次为入口位置、中间位置与出口位置。若这条匿名链路中有多个节点都由攻击者提供或有多个节点位于攻击者的流量监测范围之内，则这一攻击者可以依据多个节点通信流量之间的关系进行流量关联攻击[6~8]，关联 Tor 用户与其通信目标，破坏 Tor 网络的匿名性。

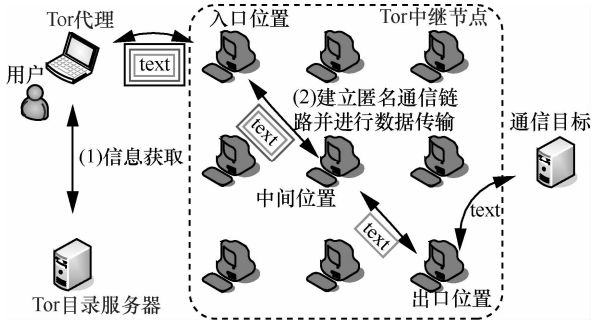


图 1　Tor 匿名通信网络的基本构成与通信原理

Tor 网络中影响其匿名性的核心算法是 Tor 的建路算法——Tor 代理以网络中各中继节点的带宽为权重，在考虑节点类型标志等因素[9]的同时，随机选择 3 个中继节点建立匿名通信链路。文献[10, 11]对这一算法进行了较为详细的分析。此外，为了保障 Tor 网络所提供的匿名性，Tor 在链路建立算法中应用了许多的限制策略以防止流量关联攻击：/16 网段限制限制 Tor 代理在建立匿名链路时使用来自同一/16 网段的 Tor 中继节点；Edman 等在文献[12]中提议将/16 网段限制扩大为自治域限制，不允许在同一链路中使用来自同一自治域的多个节点；类似地，Tor 将同一用户或机构所运营的多个节点视为一个节点家族，避免在同一链路中使用来自这一家族的多个节点[9]。虽然 Tor 家族是定义在节点与节点运营者关系之上的，但在 Tor 网络中并不存在关于这一对应关系的直接信息。在实际 Tor 网络中，Tor 家族是由 Tor 节点之间的相互声明来确定的。如果一个志愿者提供了多个 Tor 中继节点，它应当为每个中继节点配置“MyFamily”声明，将它所管理的所有 Tor 节点在“MyFamily”声明中列出；这一声明将通过 Tor 目录服务器发布给 Tor 代理供链路建立算法参考。

虽然 Tor 已经成为网络对抗和学术研究的一个热点领域，但针对 Tor 家族的研究和分析并不多见。在网络对抗方面，《卫报》刊文描述了 NSA 通过与运营商合作、Tor 指纹探测、流量中转等手段所实施的 Tor 用户追踪策略[13]。在学术研究方面，学者们已从 Tor 的匿名性增强、性能提高、规模扩展和抗审查等多个角度对 Tor 网络进行了深入的研究。在测量分析方面，Tor Metric Portal[2]项目研究了 Tor 网络中的用户及中继节点的规模与运行规律，Mccoy 等也就 Tor 节点的带宽、稳定性等进行了分析[14]。然而，由于研究动机与目的的不同，上述研究及测量分析都是针对 Tor 网络整体进行的，并未对 Tor 家族节点进行深入研究。

文献[11]将 Tor 节点分为家族节点与非家族节点 2 类，通过对比发现 Tor 家族节点在 Tor 网络性能提高与规模扩展上正扮演着的越来越重要的角色。本文将文献[11]这一分类级的测量分析工作更深入一步（如图 2 所示），基于近两年来的 Tor 网络数据，提取了 Tor 网络中由不同志愿者所运营的多个节点家族，进行了家族级的节点测量分析，主要贡献包括以下 2 个方面。

1) 基于两年来 Tor 网络数据的统计与分析全面揭示了 Tor 节点家族的诸如规模、地理分布等特性。这些分析验证了 Tor 家族设计在 Tor 网络中不可替代的地位，说明了它在防止流量关联攻击与增强匿名性方面所起到的重要作用。发现 Tor 网络中的部分超级节点家族，这些超级家族大都由一些公司、组织或社团等团体机构所运营。

2) 针对 Tor 节点家族的安全性分析说明了恶意 Tor 节点家族对 Tor 网络可用性带来的挑战，揭露了 Tor 网络中错误家族配置所造成的隐藏家族现象的普遍性及其对 Tor 网络匿名性所造成的潜在威胁。这些分析构成了 Tor 家族相关未来研究的基础与动机，为未来研究指明了方向：改进 Tor 家族设计的实现机制，研究 Tor 家族错误配置及恶意家族发现方法。

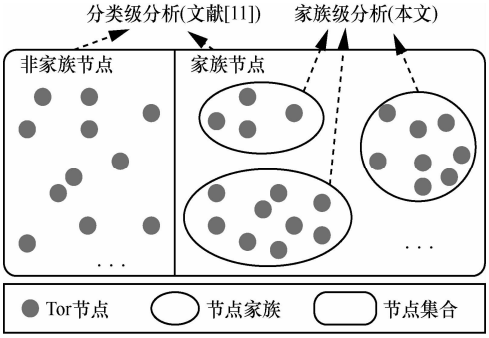


图 2 分类级/家族级 Tor 家族设计测量与分析

2 研究方法概述

2.1 方法

记 Tor 网络中的节点集合为 N，对任意节点 $n\in N$，使用 $F(n)$ 集合来表示节点 n 的家族声明，$F(n)=\{n'\in N\mid n\text{将节点}n'\text{列在了其家族声明中}\}$。如果节点 n 没有家族声明，则 $F(n)=\phi$。基于这一定义，节点 n_i 与节点 n_j 之间的关系有如下几种。

1) $n_i \rightleftharpoons n_j$，节点 n_i 与 n_j 之间存在有直接双向的家族声明，即 $n_i\in F(n_j)$ 且 $n_j\in F(n_i)$。

2) $n_i \leftrightarrow n_j$，节点 n_i 与 n_j 之间存有直接的家族声明，即 $n_i\in F(n_j)$ 或 $n_j\in F(n_i)$。

3) $n_i \sim n_j$，节点 n_i 与 n_j 之间存有直接或间接的家族声明，即存在节点 $n_{k_i},n_{k_2},\cdots,n_{k_m}\in N$ 使得 $n_i \leftrightarrow n_{k_1}$，$n_{k_1} \leftrightarrow n_{k_2}$，…，$n_{k_{m-1}} \leftrightarrow n_{k_m}$，$n_{k_m} \leftrightarrow n_j$。

可以看出，上述所定义的 $n_i \rightleftharpoons n_j$、$n_i \leftrightarrow n_j$ 与 $n_i \sim n_j$ 关系是对称的，且满足以下关系：$n_i \rightleftharpoons n_j \Rightarrow n_i \leftrightarrow n_j \Rightarrow n_i \sim n_j$。

在图 3 中，将 $n_j\in F(n_i)$ 表示为从节点 n_i 到 n_j 的一条有向边。图中的 n_1、n_2 和 n_3 之间有着双向直接的家族声明，即 $n_1 \rightleftharpoons n_2$、$n_2 \rightleftharpoons n_3$ 且 $n_3 \rightleftharpoons n_1$。另外，$n_4$ 的家族声明中列出了 n_5 与 n_6，即 $F(n_4)=\{n_5,n_6\}$，可表示为 $n_4 \leftrightarrow n_5$ 与 $n_4 \leftrightarrow n_6$。虽然 n_5 与 n_6 之间没有直接的家族声明关系，但它们都可以通过节点 n_4 相连，记为 $n_5 \sim n_6$。图中的节点 n_7 与 n_8 没有家族声明，也没有被包含在其他节点的家族声明中。

文献[15]从修正家族声明的角度出发，将节点之间的“$\sim$”关系修正为“$\rightleftharpoons$”关系，并基于“$\rightleftharpoons$”关系给出了 $F^{(1)}$ 家族的定义。依据相同的思想，本文简化上述过程，直接将 Tor 节点家族定义在“$\sim$”关系之上——对于 Tor 节点子集 S（$S\subseteq N$）而言，若它满足如下条件，则将 S 视为一个节点家族。

1) $|S|\geqslant 2$，节点家族中的节点数目不小于 2。

2) 对于任意的 $n_i,n_j\in S$，$n_i \sim n_j$ 都成立，节点家族内的任意 2 个节点之间都存在着直接或间接、双向或单身的家族声明。

3）对于任意的 $n_i\notin S$ 与任意的 $n_j\in S$，$n_i \sim n_j$ 不成立。

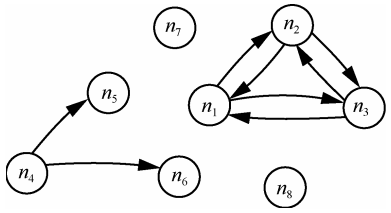


图 3 将 Tor 的家族声明视为一个顶点—连边

上述对 Tor 节点家族的定义与文献[15]中的 $F^{(1)}$ 家族一致，都将节点之间直接或间接、单向或双向的家族声明视为有效声明，可以更好地反应 Tor 家族特性，为后续的测量分析工作提供可靠的数据基础。另外，由于上述定义直接定义在“$\sim$”关系之上，本文实验可以更方便地借助顶点—连边图相关的通用算法来提取 Tor 节点家族——每个节点家族都可以被看作节点声明图中的一个连通子集。在图 3 中，$F_1=\{n_1,n_2,n_3\}$ 与 $F_2=\{n_4,n_5,n_6\}$ 是 2 个节点家族，而 n_7 与 n_8 不属于任何一个节点家族。

2.2 数据与实验

本文实验数据取自 2011 年 1 月至 2012 年 12 月的 Tor 网络运行真实数据，截取了 24 组 Tor 网络数据，均匀分布在每个月 20 日的每个小时。具体而言，在 2011 年 1 月 20 日选取 00:00 的 Tor 网络数据进行分析，在 2011 年 2 月 20 日选取 01:00 的 Tor 网络数据进行分析…在 2012 年 12 月 20 日选取 23:00 的 Tor 网络数据进行分析。

在这 24 组数据中，每一组都包括当时 Tor 网络的节点快照及节点快照中各节点所对应的节点描述符。这些数据所包含的信息与 Tor 代理在当时所能获取的信息一致，因而基于这些数据的分析能够反应 Tor 代理的视角；同时，这些数据包含了当时 Tor 网络中所有的在线中继节点，使得可以从全局层面出发来挖掘 Tor 网络的家族特性。另外，选取均匀分布于两年内的 24 组数据进行分析，也使得本文的结果更具一般性。

在每组数据中，记 Tor 网络中的在线节点集合为

N_c。对于每个节点$n \in N_c$，都从它的描述符文件中提取其家族声明$F(n)$。所有在线节点与家族声明中所列举的节点（$\bigcup_{n\in N_c} F(n)$）构成了该组数据的节点全集$N = N_c \bigcup \left(\bigcup_{n\in N_c} F(n) \right)$。在这 24 组数据中，集合$N$平均包含 2 794 个节点，其中在线节点约 2 647 个，另外的 147 个节点虽然被列在了节点的家族声明中，但在截取 Tor 网络快照时，它们并不在线。

举例而言，在 2012 年 12 月 20 日 23:00 的节点全集N与它们之间的节点声明所构成的顶点—连边如图 4 所示。使用连通子集发现算法来提取 Tor 节点家族。在这组数据的节点集合N中，实验共发现了 173 个节点家族，这些节点家族包含有 482 个在线节点和 209 个离线节点。记从每组数据的节点全集N中所提取的 Tor 节点家族为$N \cdot family$，实验从 24 组数据中共提取了 3 390 个$N \cdot family$，平均每组数据 141 个。

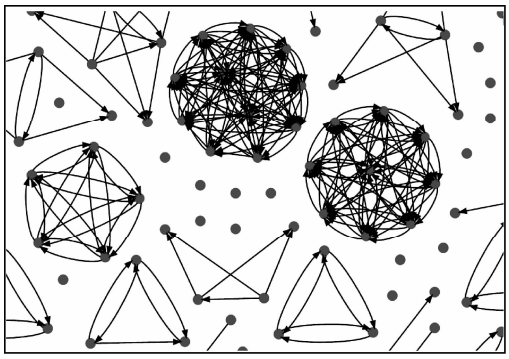


图 4　实际 Tor 网络的节点及节点的家族声明

对于离线节点而言，并不能获取其详细信息，因而也在每组数据的在线节点集合N_c上进行了上述节点家族的发现实验。这一实验忽略了所有的离线节点以及指向这些离线节点的家族声明。在 2012 年 12 月 20 日 23:00 的在线节点集合N_c上的家族声明为图 4 的一个子图，仅包含了在线的节点及它们之间连边。从本组N_c中发现了 102 个节点家族，这些节点家族包含了 409 个在线节点。记从每组数据的在线节点集合N_c中所提取的 Tor 节点家族为$N_c \cdot family$，实验从 24 组数据中共提取了 2 023 个$N_c \cdot family$，平均每组数据 84 个。

下文的分析测量基于上述实验中所提取的$N \cdot family$与$N_c \cdot family$，$N \cdot family$主要用于揭示 Tor 节点家族的规模、在线比例等规律，而$N_c \cdot family$则用于分析诸如节点家族带宽、地理分布等需要详细信息支持的家族特性。为了使实验结果更具一般性，下文实验中所展示的 Tor 网络节点家族特性都是两年内 24 组数据的平均结果。

3　家族级的测量分析

3.1　节点家族的规模

图 5 给出了 Tor 网络中各节点家族的规模分布。所发现的最大$N \cdot family$包含有 44 个 Tor 节点，这些节点的昵称都以“PPrivCom”为前缀，它们是由 perfect-privacy 组织所提供的，本文将这一节点家族称为 perfect-privacy 家族。在实验进行时，这 44 个节点中最多有 42 个同时在线，是实验中所发现的最大的$N_c \cdot family$。从图 5 中可以得出如下结论。

1) 大多数的节点家族都很小，大约有 90%的节点家族都由 5 个以下的 Tor 节点构成。实际上，实验发现的$N \cdot family$与$N_c \cdot family$的平均规模分别为 3.64 和 3.66。

2) $N_c \cdot family$的总数小于$N \cdot family$。一个节点家族中的节点并不一定总是同时在线的，仅有一个节点在线的$N \cdot family$不符合 2.1 节中对节点家族的定义，因而不会被认为是一个有效的$N_c \cdot family$。

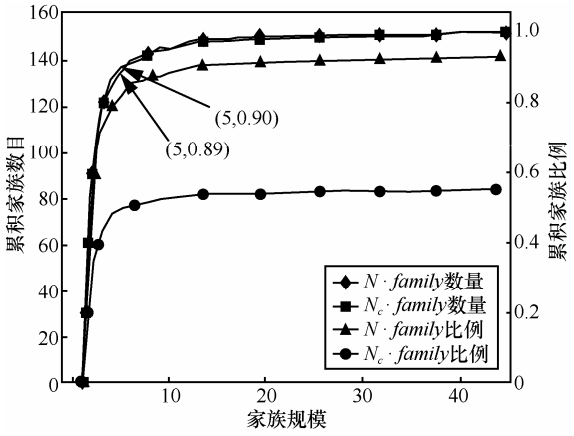


图 5　Tor 网络中节点家族的规模分布

3.2　节点在线比例

不同规模的*Nfamily*中的在线节点比例如图 6 所示。在这一分析中，依据家族规模（2，3，4，5-8，9+）将实验发现的*Nfamily*分为了 5 类。每一类*Nfamily*所包含的在线节点数目与这一类节点总数的比较如图 6 所示。从图中可以看出，在一个节点家族中，平均约有 65%的节点会同时在线，而节点在线比例与其家族规模并无明显关联。

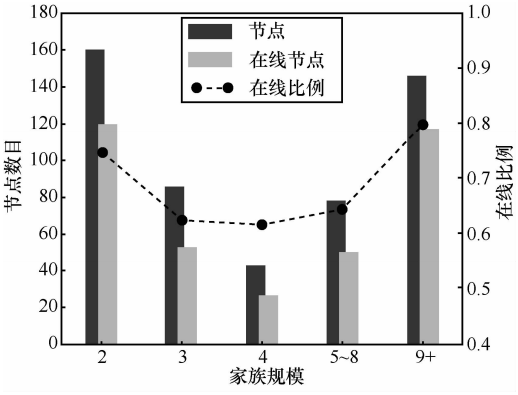


图 6 不同规模的节点家族中在线节点数目及比例

3.3 节点家族的带宽分布

$N_c \cdot family$ 所提供的带宽分布如图 7 所示。这些节点家族所提供的带宽分布极广，从 4 kbit/s 到 1.8 Gbit/s 不等。大约有 70%的节点家族所提供的带宽小于 9.4 Mbit/s。另外，大约有 1.5%的节点家族提供了 500 Mbit/s 以上的带宽。正如文献[11]所发现的一样，这些节点家族的高带宽为 Tor 网络的性能与规模做出了极大的贡献，然而，一个拥有着高带宽的 Tor 节点家族也可能会引起攻击者的极大关注。

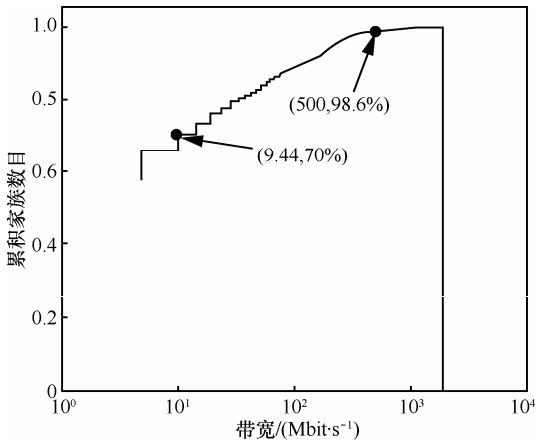


图 7 节点家族的带宽分布

3.4 节点家族的地理分布

表 1 给出了 $N_c \cdot family$ 中节点的地理分布情况。在实验中，检查了节点家族中的各个节点所在的/16 子网、自治域及国家归属信息。基于这一分析，实验发现了一些跨/16 子网、跨自治域甚至跨国的节点家族。例如：上文所提到的 perfect-privacy 节点家族中的节点分布在 27 个不同的/16 子网、23 个不同的自治域及 17 个不同的国家中，它是一个跨网、跨自治域、跨国的节点家族。

表 1 Tor 节点家族内节点的地理分布

节点家族	家族数目	家族比例
跨/16 子网家族	66	79%
跨自治域家族	56	67%
跨国家族	29	35%

从表 1 可以看出，79%的节点家族都是跨/16 子网的家族。如果没有 Tor 家族设计，目前 Tor 建路算法所实行的/16 子网限制根本不能防止 Tor 在同一匿名链路中使用来自同一 Tor 家族的多个节点，会严重影响 Tor 网络的匿名性。在实验中，有 67%及 35%的 Tor 节点家族分别是跨越自治域、跨越国家的。因而，Matthew 等人提出的基于自治域限制、甚至基于国家限制的建路算法都不能排除来自 Tor 家族的节点。可见，Tor 家族是 Tor 网络的一个极为重要的设计，在保障了 Tor 匿名性方面发挥了不可替代的作用。

3.5 节点家族运营者

实验中，有多个 $N_c \cdot family$ 的规模超过 10，检查这些节点家族中各节点的联系人、whois 等信息，挖掘部分家族的运营者身份，如表 2 所示。大部分的超级节点家族都由一些公司、组织或社团等团体机构所提供。这些团体机构充足的财力及技术水平也许可以用于解释文献[11]所发现的现象：相对于其他节点而言，Tor 节点家族中的节点有着更高的带宽及稳定性。

表 2 超级家族的提供者

家族	运营者	详细信息
PPrivCom	perfect-privacy.com	公司，致力于加密与网络匿名通信
torservers	torservers.net	组织，运行高带宽 Tor 节点的组织
DFRI	dfri.se	组织，致力于内容数字版权加密保护技术的组织
chaoscomputerclub	ccc.de	社团，欧洲最大的黑客社团
spfTOR	privacyfondation.de	基金，德国隐私保护基金
Torrorists	torrorists.de	社团，联合运行 Tor 节点的志愿者社团

4 节点家族安全性

4.1 基于节点家族的“网桥收割”攻击

本节讨论基于 Tor 节点家族的网桥收割攻击，说明恶意 Tor 节点家族给 Tor 网络所带来的可用性威胁。

Tor 网桥[3]是 Tor 网络可用性与抗审查能力的一

个重要保障。与普通中继节点类似，网桥节点可以被用在 Tor 匿名链路的入口位置。然而，与普通的 Tor 中继节点不同的是，Tor 网桥节点通过带外的如 HTTPS、email、社交网络等方式发布，用户或攻击者不能得到 Tor 网桥节点的完整列表，因而难以阻止用户与网桥之间的连接。在实际中，除了从 Tor 官方发布渠道枚举 Tor 网桥外，还存在另一种收集 Tor 网桥的攻击方式：网桥收割（bridge harvest）攻击。通过利用恶意的 Tor 中间节点（链路中的第二跳），攻击者可以收集连接到它的第一跳节点的信息并从中筛选 Tor 网桥节点。目前已有一些研究[10, 16, 17]实现了这种攻击并证明了其有效性。

如文献[11]所示，较其他节点而言，来自 Tor 节点家族的节点提供了更高的带宽与稳定性。另外，由于节点家族中的节点都来自于同一运营者，这些节点及其所运行的操作系统的漏洞、甚至管理员密码都可能相似或相同，攻击一个节点家族所付出的代价可能与攻击此节点家族中的一个 Tor 节点的代价类似。Tor 节点家族这种高带宽、集中控制的特点使它们成为了一个理想的攻击目标，吸引攻击者利用它们来进行有效的网桥收割攻击。

本文使用捕获率[18] $P_c(F_i)$ 来衡量一个 Tor 代理使用某 Tor 网桥建立的链路经过攻击者所控制的节点家族 F_i 中节点的概率。以 $P(p,n)$ 表示节点 n 被 Tor 客户端选中用于构建匿名链路中 p 位置的概率[11]，则 Tor 代理将来自 Tor 节点家族 F_i 中的节点选作一条链路中的中间节点概率（即捕获率）表示为

$$P_c(F_i)=\sum_{n_j\in F_i}P(p_m,n_j) \tag{1}$$

Tor 用户通过某网桥建立 k 次链路后，攻击者捕获此网桥的概率可表示为

$$P_c(F_i,k)=1-\left(1-P_c(F_i)\right)^k \tag{2}$$

本文考虑 Tor 网络中最大的节点家族（perfect-privacy 家族）被攻击者利用来进行网桥收割攻击时的情况。为了提高攻击效率，攻击可以如文献[10]所述，合理配置并控制这一家族中节点的运行以使之获得“middle”标志，以更大的概率被选为链路的中间节点。以 2012 年 12 月 20 日 23:00 的 Tor 网络为例，网络中有 36 个 perfect-privacy 家族节点，结合 Tor 网络快照中的带宽权重信息，由式(1)可得到网桥捕获率为 $P_c(F_{\text{perfect-privacy}})=0.046$，即此网桥第一次被使用就会被攻击者捕获的概率为 4.46%。当这一网桥被使用 48 次后，攻击者发现它的概率将超过 90%。

由此例可以看出，恶意的 Tor 节点家族可以实现高效的网桥收割攻击，给 Tor 网络的抗审查设计带来了巨大的威胁。这一分析也警示 Tor 家族节点的运营者们提高其系统安全性，防止攻击者利用这些节点并借由它们发起其他攻击。

4.2　Tor 网络中的隐藏家族现象

虽然 Tor 家族是定义在节点与节点运营者关系之上的，但目前的 Tor 版本都依赖各个中继节点间的家族声明来实现其家族设计。其设计与实现上的不一致性导致了 Tor 家族机制对 Tor 节点运营者的严重依赖，也可能会导致隐藏家族现象：如果一个提供了多个 Tor 中继节点的志愿者忘记或错误配置了“MyFamily”声明，这些中继节点将不会被 Tor 代理视为同一家族，构成了一个隐藏的节点家族，不能避免 Tor 代理在同一链路中使用它们，从而降低了 Tor 用户的匿名性。基于两年来的 24 组数据，本节对 Tor 网络节点及运营者数目进行了对比，揭示了隐藏家族现象的普遍性及其对 Tor 用户的匿名性所构成的潜在威胁。

在本文所收集的每组在线集合 N_c 中，将不属于任意一个 $N_c\cdot family$ 的中继节点称为非家族节点。举例而言，2012 年 12 月 20 日 23:00 的 Tor 网络共包含 2 962 个节点，从中发现了 102 个节点家族（包含 409 节点），不被包含在任意一个节点家族中的 2 553 个节点被称为非家族节点。在这些非家族节点中有 1 328 个节点都设置了运营者联系邮箱，邮箱数目去重后共计 1 260 个，低于节点数目。实验将运营者邮箱作为运营者的唯一标识，如图 8 所示，在其他的 23 组数据中也发现了类似的现象，设置了联系邮箱的非家族节点数目与运营者总数并不一致，这一差异说明，在这些非家族节点中，仍存在有同一运营者提供多个节点的现象，志愿者忘记或错误配置“MyFamily”声明的现象较为普遍。

图 9 揭示了由于错误的家族配置所导致的隐藏家族现象的普遍性。在实验所覆盖的 24 组数据中都存在隐藏家族，平均每组数据中 27 个；相应地，2.2 节的实验从每组数据中发现的 $N_c\cdot family$ 平均为 84 个。由此可见，错误配置所引起的隐藏家族现象在 Tor 网络中极为普遍性。举例而言，实验中

发现的联系方式为“Privacy Republic < tor-nodes AT privacyrepublic dotorg >”的志愿者在 2012 年 9 月 20 日 20:00 提供了 14 个 Tor 中继节点，都没有按照 Tor 家族设计的要求设置“MyFamily”声明，因而构成了一个隐藏的 Tor 节点家族，不能被 Tor 代理所识别。

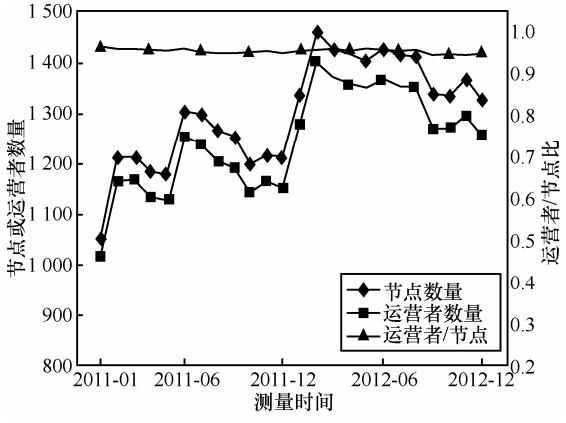


图 8 非家族节点的数目与其运营者数目对比

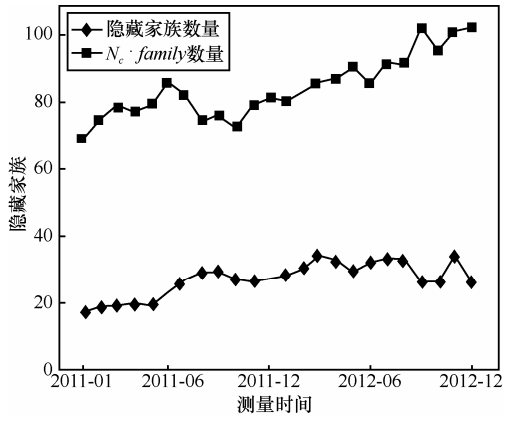


图 9 隐藏家族现象的普遍性

另外，实验还分析了 Tor 隐藏家族对 Tor 匿名性的影响。由于隐藏家族不能被 Tor 代理所识别，Tor 代理可能会选择来自同一隐藏家族的多个节点建立匿名链路，将这种链路称为低匿名链路，以低匿名链路发生的概率为标准来衡量隐藏家族对 Tor 匿名性所造成的影响。实验模拟了 Tor 链路建立算法，首先选择链路的出口位置节点，再在遵从家族限制的前提下，依次选择入口位置及中间位置的节点构建匿名链路。针对每组数据，实验都模拟建立了 10 万条匿名通信链路，从每组实验中发现低匿名链路平均数目为 969 条，即平均有 0.97% 的链路都受到了隐藏家族的影响。考虑到 Tor 网络中巨大的用户基数及每个用户所使用链路的频繁更新，隐藏家族现象的普遍性对 Tor 匿名性的影响不容忽视。

5 结束语

基于近两年来的 24 组 Tor 网络运行数据，本文对 Tor 节点进行了家族级的测量分析，揭示了 Tor 节点家族的特点与运行规律。分析结果验证了 Tor 家族设计在保障 Tor 网络匿名性中所发挥不可替代的重要作用。另外，还说明了 Tor 节点家族的误用对 Tor 网络带来的威胁，揭示了 Tor 网络中隐藏家族现象的普遍性。

目前，正在将此研究深入到节点级——研究 Tor 节点家族内部的各个节点之间的关系与行为特点。希望能够基于这一研究发现潜在的 Tor 节点家族：1) 发现 Tor 家族节点的错误配置及隐藏家族情况，帮助运营者们修正错误的家族配置；2) 区分普通的 Tor 节点家族及可疑恶意节点家族，避免建路过程中使用来自恶意节点家族的 Tor 中继节点。

参考文献：

[1] DINGLEDINE R, MATHEWSON N, SYVERSON P. Tor: the second-generation onion router[A]. Proceedings of the 13th Conference on USENIX Security Symposium[C]. 2004.

[2] Tor metrics portal[EB/OL]. https://metrics. torproject.org/:The Tor Project, 2013.

[3] DINGLEDINE R, MATHEWSON N. Design of a blocking-resistant anonymity system[R]. The Tor Project, 2006.

[4] 高峰, 杨明, 罗军舟等. Tor 匿名通信流量在线识别方法[J]. 软件学报, 2013, 24(3): 540-556
HE G, YANG M, LUO J Z, *et al*. Online identification of Tor anonymous communication traffic[J]. Journal of Software, 2013, 24(3): 540-556.

[5] THE TOR PROJECT. Tor directory protocol, version 3[EB/OL]. https://gitweb.torproject.org/torspec.git/blob/HEAD:/dir-spec.txt, 2013.

[6] FEAMSTER N, DINGLEDINE R. Location diversity in anonymity networks[A]. Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society[C]. 2004. 66-76.

[7] STEVEN J, MURDOCH, PIOTR ZIELIŃSKI. Sampled traffic analysis by internet-exchange-level adversaries[A]. Proceedings of the 7th International Conference on Privacy Enhancing Technologies[C]. 2007.167-183.

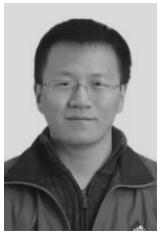
[8] BAUER K, MCCOY D, GRUNWALD D, *et al.* Low-resource routing attacks against tor[A]. Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society[C]. 2007.11-20.

[9] DINGLEDINE R, MATHEWSON N. Tor path specification[EB/OL]. https://gitweb.torproject.org/torspec.git?a=blob_plain;hb=HEAD;f=path-spec.txt, 2013.

[10] LING Z, LUO J, YU W, *et al*. Extensive analysis and large-scale empirical evaluation of Tor bridge discovery[A]. IEEE INFOCOM[C]. 2012.2381-2389.

[11] WANG X, SHI J, FANG B, *et al*. An empirical analysis of family in the Tor network[A]. IEEE International Conference on Communications[C]. Budapest, 2013.1995-2000.
[12] EDMAN M, SYVERSON P. As-awareness in tor path selection[A]. Proceedings of the 16th ACM conference on Computer and Communications Security[C]. 2009.380-389.
[13] BRUCE SCHNEIER. Attacking Tor: how the NSA targets users' online anonymity[EB/OL]. http://www.theguardian.com/world/ 2013/oct/04/tor-attacks-nsa-users-online-anonymity: Guardian News and Media Limited, 2013.
[14] MCCOY D, BAUER K, GRUNWALD D, *et al*. Shining light in dark places: understanding the tor network[A]. Proceedings of the 8th International Symposium on Privacy Enhancing Technologies[C]. Berlin, 2008. 63-76.
[15] WANG X, SHI J, GUO L. Towards analyzing family misconfiguration in tor network[A]. Computer Science and its Applications, Lecture Notes in Electrical Engineering[C]. 2012.503-514.
[16] VASSERMAN E, JANSEN R, TYRA J, *et al*. Membership-concealing overlay networks[A]. Proceedings of the 16th ACM Conference on Computer and Communications security[C]. 2009. 390-399.
[17] MCLACHLAN J, HOPPER N. On the risks of serving whenever you surf: vulnerabilities in Tor's blocking resistance design[A]. Proceedings of the 8th ACM Workshop on Privacy in the Electronic Society[C]. 2009.31-40.
[18] LI C, XUE Y, DONG Y, *et al*. Super nodes in Tor: existence and security implication[A]. Proceedings of the 27th Annual Computer Security Applications Conference[C]. 2011.217-226.

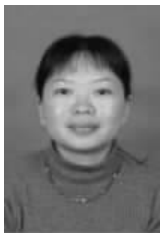
作者简介：



王啸（1986-），男，河北邢台人，中国科学院博士生，国家计算机网络应急技术处理协调中心工程师，主要研究方向为匿名通信、隐蔽通信与信息对抗。

方滨兴（1960-），男，江西万年人，中国工程院院士，主要研究方向为计算机网络和信息安全。

刘培朋（1987-），男，山东日照人，中国科学院博士生，主要研究方向为匿名通信、隐蔽通信和信息对抗。



郭莉（1969-），女，湖南株洲人，中国科学院高级工程师、教授、博士生导师，主要研究方向为信息内容安全管理、网络安全、数据流处理和网络流计算等。

时金桥[通信作者]（1978-），男，黑龙江哈尔滨人，博士，中国科学院副研究员，主要研究方向为信息对抗技术和隐私保护技术。E-mail: shijingqiao@iie.ac.cn。