北京郵電大學
Beijing University of Posts and Telecommunications

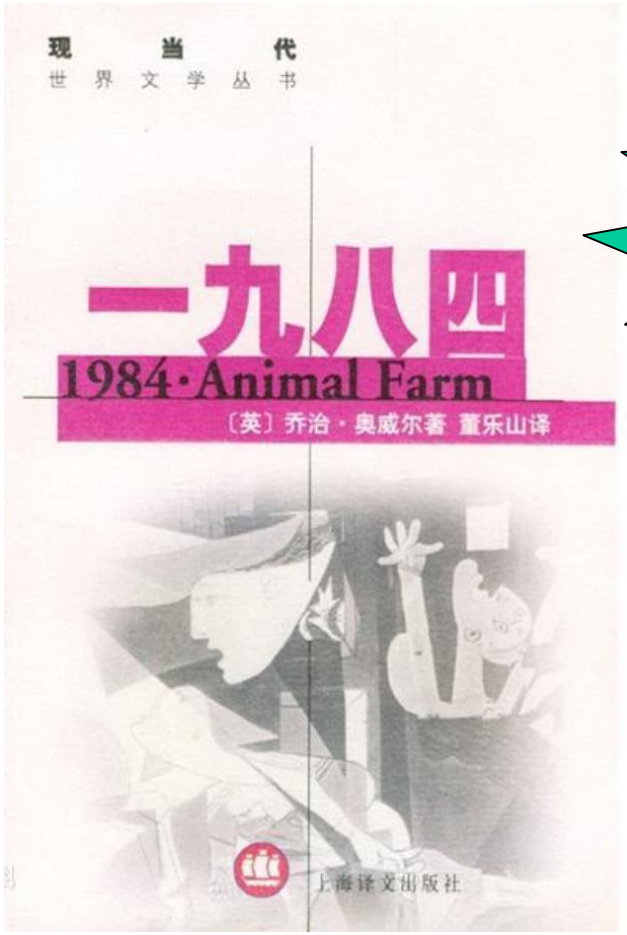
大数据安全与隐私保护
匿名通信技术

石瑞生
网络空间安全学院

提纲

- 简史
- 概念
- 原理
- 应用系统

匿名通信技术简史



乔治·奥威尔（1949）

“老大哥”
在看着你！

David Chaum

American computer scientist

David Lee Chaum is an American computer scientist and cryptographer. He is known as a pioneer in cryptography and privacy-preserving technologies, and widely recognized as the inventor of digital cash. He is Jewish and was born to a Jewish family in Los Angeles. Wikipedia

Born: January 1, 1955 (age 65 years)

Education: University of California, Berkeley

Books: Advances in Cryptology

Known for: DigiCash, Ecash, MORE

美国密码学家 David Chaum 在 1985 年说：Mix Network 可以用于匿名网络通讯、匿名电子货币等多种用途，使老大哥完全过时。

Chaum, D., Security Without Identification: Transaction Systems to Make Big Brother Obsolete, Communications of the ACM, vol. 28, no. 10, October 1985, pp. 1030-1044.

基础理论：问题与方法

1979年，David Chaum在他的硕士论文中提出了混合网络的概念，来解决通信中元数据泄漏导致的隐私问题。

大多数的安全保护措施都集中在防止窃听上。但是**加密的信息仍然可以被追踪，揭示谁在和谁说话**。这种跟踪称为流量分析，可能会泄露敏感信息。

- 例如，公司间协作的存在可能是机密的。
- 类似地，电子邮件用户可能不希望向世界其他地方透露他们正在与谁通信。

由于2013年斯诺登(Snowden)的爆料，**防范大规模监控**的工具需求成为了主流关注点。Tor不仅对斯诺登的泄密起到了帮助作用，而且文件的内容也支持了当时Tor不能被破解的保证。

- **计算匿名**：MIX-Net
- **信息理论匿名**：DC-Net

[1] Chaum, D., Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms, Communications of the ACM, vol. 24, no. 2, February 1981, pp. 84-88.

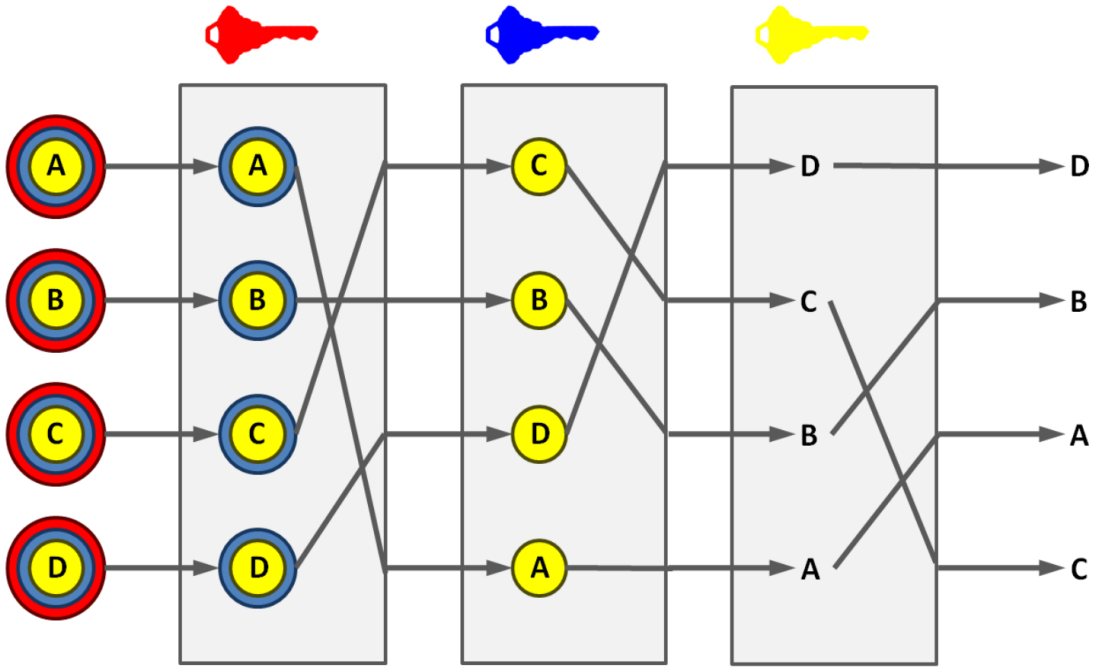
[2] Chaum, D., Security Without Identification: Transaction Systems to Make Big Brother Obsolete, Communications of the ACM, vol. 28, no. 10, October 1985, pp. 1030-1044.

[3] David Chaum (1988). "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability". Journal of Cryptology. 1 (1): 65–75.

MIX-Net

混合网络的概念最早是由David Chaum在1981年提出的。

基于此概念的应用程序包括匿名remailers(如Mixmaster)、洋葱路由、大蒜路由和基于密钥的路由(包括Tor、I2P和Freenet)。



简单的解密混合网。消息是在一组公钥下加密的。每个mix节点使用自己的私钥删除一层加密。节点打乱消息顺序，并将结果传输到下一个节点。

混合网络是一种路由协议,它通过使用一个代理服务器链（代理服务器被称为mix）提供了难以追踪的通信服务。

Mix接收多个发送者的消息，将消息进行重组（shuffle),以随机的顺序把这些消息发送到下一个目的地(可能是另一个混合节点)。这打破了请求的源和目的地之间的联系,使窃听者更难跟踪端到端通信。此外,Mix只知道它收到消息的上一节点,和发送重组后消息下一个目的地,使网络抵抗恶意节点。

How it works

Participant *A* prepares a message for delivery to participant *B* by appending a random value R to the message, sealing it with the addressee's public key K_b, appending B's address, and then sealing the result with the mix's public key K_m. M opens it with his private key, now he knows B's address, and he sends $K_b(message, R)$ to B.

Message format $K_m(R1, K_b(R0, message), B) \longrightarrow (K_b(R0, message), B)$

Return addresses： What is needed now is a way for *B* to respond to *A* while still keeping the identity of *A* secret from *B*.

The message from $A \longrightarrow B$:

$$K_m(R1, K_b(R0, message, K_m(S1, A), K_x), B) \longrightarrow K_b(R0, message, K_m(S1, A), K_x)$$

Reply message from $B \longrightarrow A$:

$$K_m(S1, A), K_x(S0, response) \longrightarrow A, S1(K_x(S0, response))$$

Where: K_b = B s public key, K_m = the mix's public key.

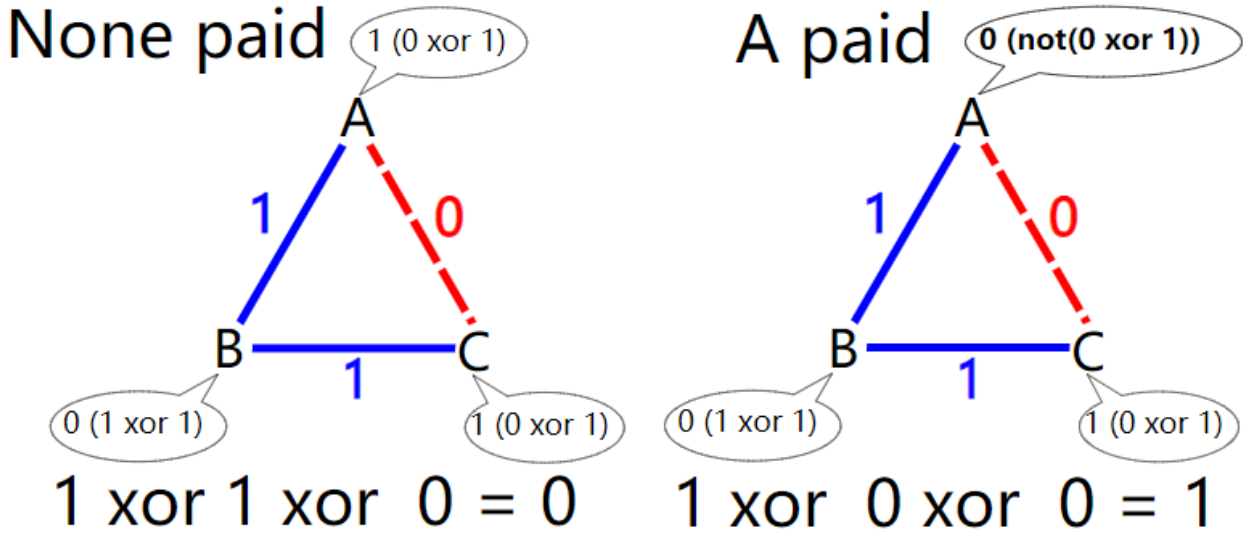
Vulnerabilities

Although mix networks provide security even if an adversary is able to view the entire path, mixing is not absolutely perfect. Adversaries can provide **long term correlation attacks** and track the sender and receiver of the packets.

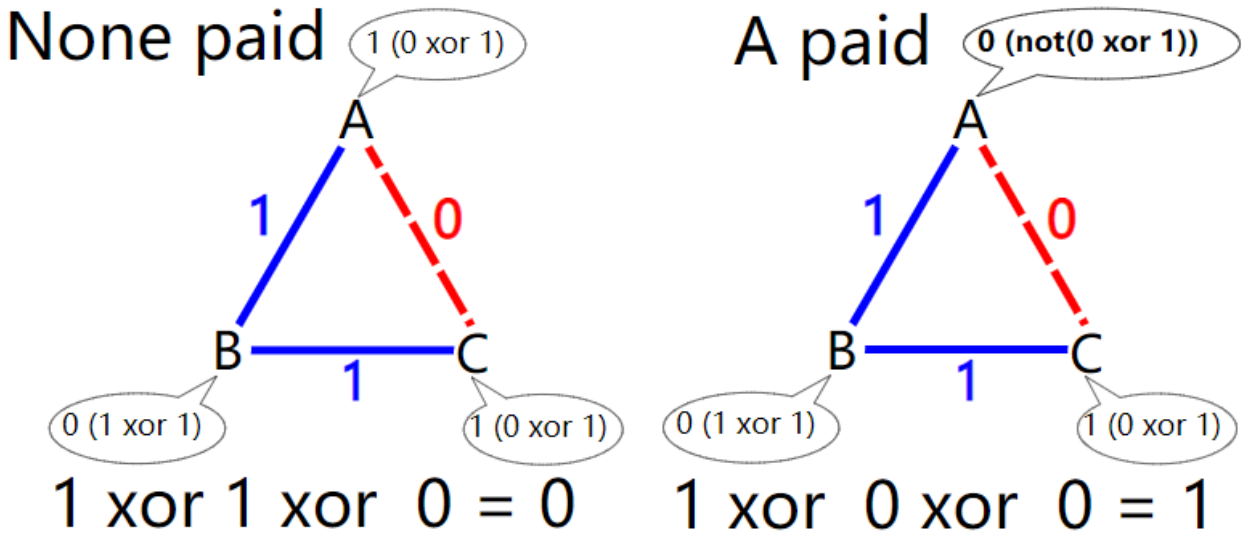
密码学家的晚餐 - DINING CRYPTOGRAPHER'S PROBLEM

我知道是你们中的某个人付了钱，但是无法知道具体是谁。

Three cryptographers gather around a table for dinner. The waiter informs them that the meal has been paid for by someone, who could be one of the cryptographers or the National Security Agency (NSA). The cryptographers respect each other's right to make an anonymous payment, but want to find out whether the NSA paid. So they decide to execute a two-stage protocol.



算法示例



第一阶段:
假设在抛硬币之后，A和B共享一个秘密位1,A和C共享0,B和C共享1。

在第二阶段，每个密码学家公开宣布一个bit，即:
如果他没有付这顿饭的钱，公布左右两边的投掷硬币结果的异或（XOR）值。
如果他付了餐费，就公布（与异或）相反的值。

一个简单地计算来得到最终的结果：三位的异或。
如果结果是0，则意味着没有任何密码破译人员为此付费(因此，NSA为此买单)。
否则，其中一个密码学家付钱，但其他密码学家仍然不知道他们的身份。

DC网络很容易推广为允许每轮传输多于一个比特，大于三个参与者的组以及除二进制数字0和1之外的任意“字母”。

课后阅读材料

- 选择一篇论文阅读

^ a b David Isaac Wolinsky; Henry Corrigan-Gibbs; Bryan Ford; Aaron Johnson (October 8–10, 2012). *Dissent in Numbers: Making Strong Anonymity Scale*. 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI). Hollywood, CA, USA.

^ Philippe Golle; Ari Juels (May 2–6, 2004). *Dining Cryptographers Revisited* (PDF). Eurocrypt 2004. Interlaken, Switzerland.

^ Franck, Christian (2008). *New Directions for Dining Cryptographers* (PDF) (M.Sc. thesis).

^ a b Henry Corrigan-Gibbs; David Isaac Wolinsky; Bryan Ford (August 14–16, 2013). *Proactively Accountable Anonymous Messaging in Verdict*. 22nd USENIX Security Symposium. Washington, DC, USA.

^ a b Henry Corrigan-Gibbs; Bryan Ford (October 2010). *Dissent: Accountable Group Anonymity*. 17th ACM Conference on Computer and Communications Security (CCS). Chicago, IL, USA. Archived from the original on 2012-11-29. Retrieved 2012-09-09.

^ Emin Gün Sirer; Sharad Goel; Mark Robson; Doğan Engin (September 19–22, 2004). *Eluding Carnivores: File Sharing with Strong Anonymity* (PDF). ACM SIGOPS European workshop. Leuven, Belgium.

^ Nikita Borisov; George Danezis; Prateek Mittal; Parisa Tabriz (October 2007). *Denial of Service or Denial of Security? How Attacks on Reliability can Compromise Anonymity* (PDF). ACM Conference on Computer and Communications Security (CCS). Alexandria, VA, USA.

应用系统

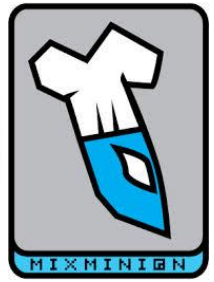
- 匿名代理、匿名存储、匿名发布、匿名邮件

洋葱路由系统
（1995年）

Web匿名代理（1996年）

匿名Web浏览
（1997年）

分布式的匿名数据存储（1999年）

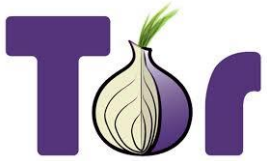


匿名邮件系统
（2002年）

匿名代理（2000年）

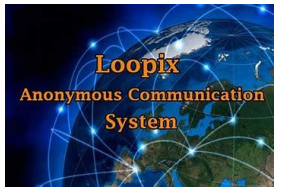
P2P模式实现文件匿名存储和检索
（2001年）

分布式P2P匿名通信系统（2003年）



第二代洋葱路由匿名通信系统
（2003年）

分布式Web服务
Zeronet
（2015年）



中延迟服务
IM、邮件
Loopix
（2017年）

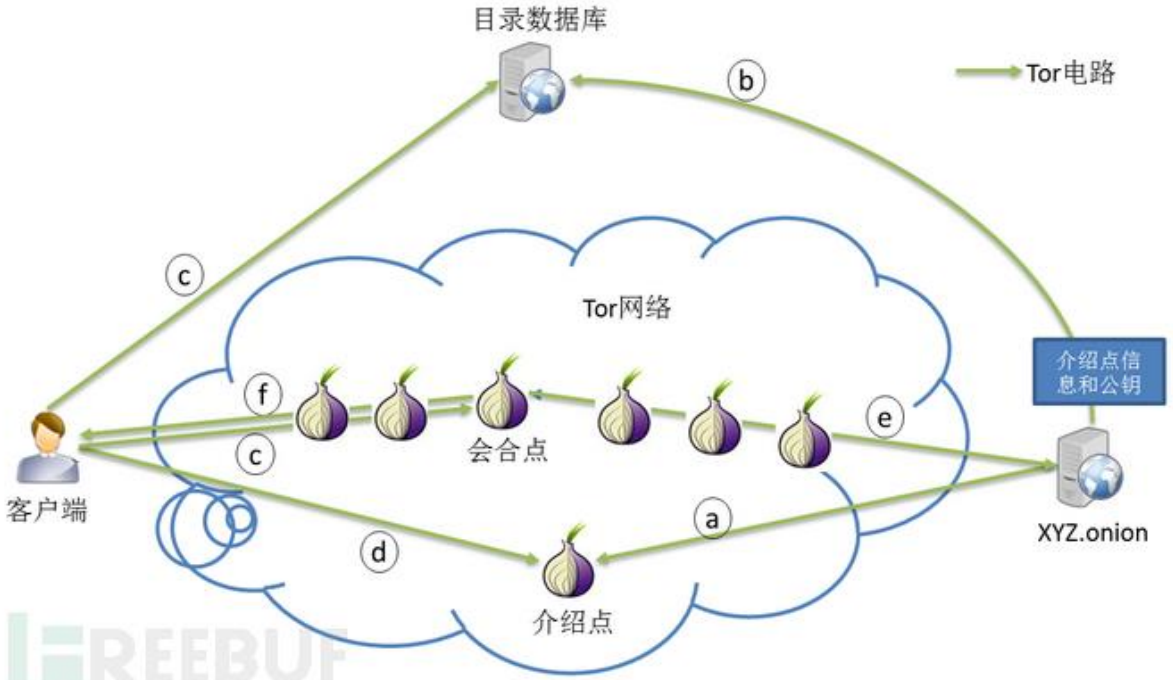
TOR

第二代洋葱路由
匿名通信系统
（2003年）

Paul Syverson

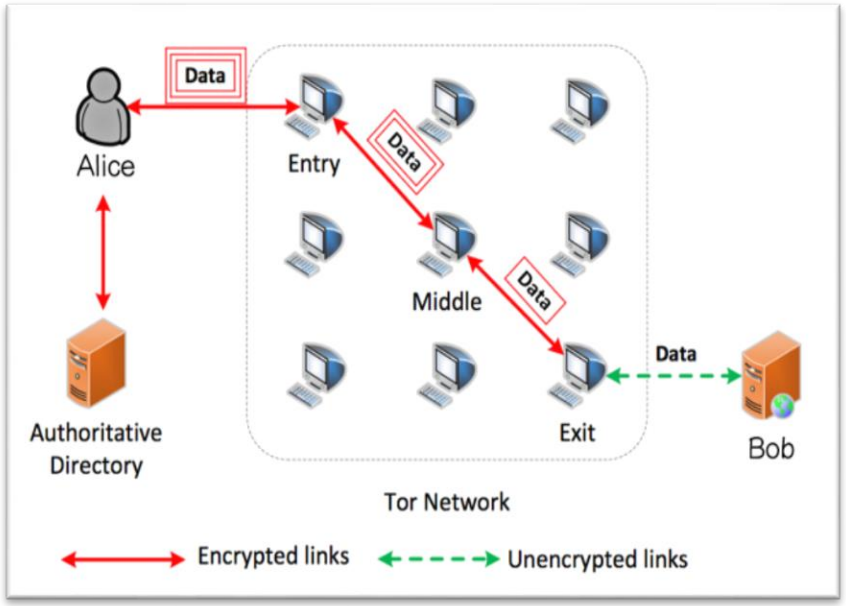
提供两套目录服务（权威目录服务和HS目录服务），分别维护节点命名空间和服务命名空间，权威目录服务可信，HS目录服务不可信

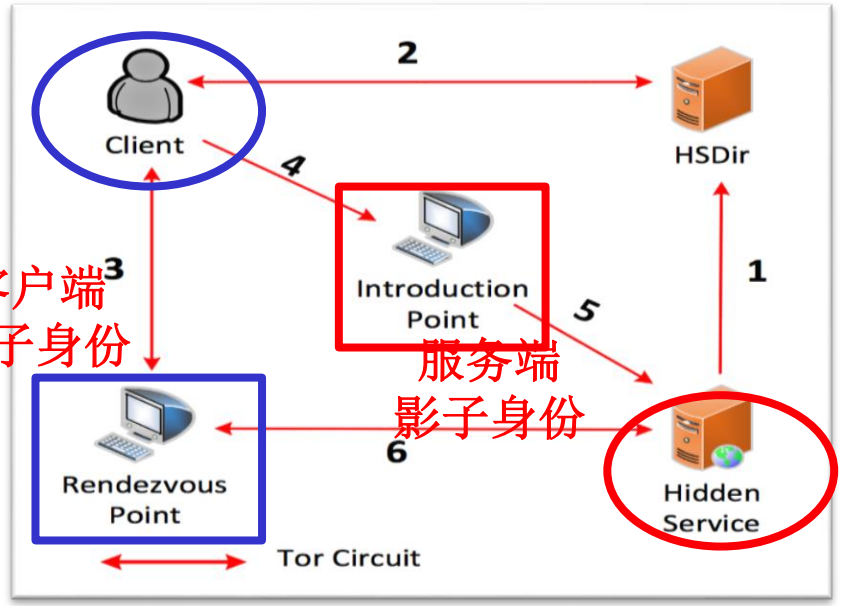
在20世纪90年代，互联网缺乏安全性以及用于跟踪和监视的能力变得越来越明显，1995年，美国海军研究实验室（NRL）的 David Goldschlag，Mike Reed 和 Paul Syverson 问自己是否有一种方法可以创建互联网连接但不会泄露谁与谁交谈，甚至不会泄漏给监控网络的人。他们给的答案就是创造和部署了首个洋葱路由研究设计方案和原型。
洋葱路由的目标是为了有一种尽可能隐私的方式来使用互联网，其想法是让流量路由通过多个服务器并在每一步进行加密。 这仍是一种对今天的 Tor 如何工作的简单解释。

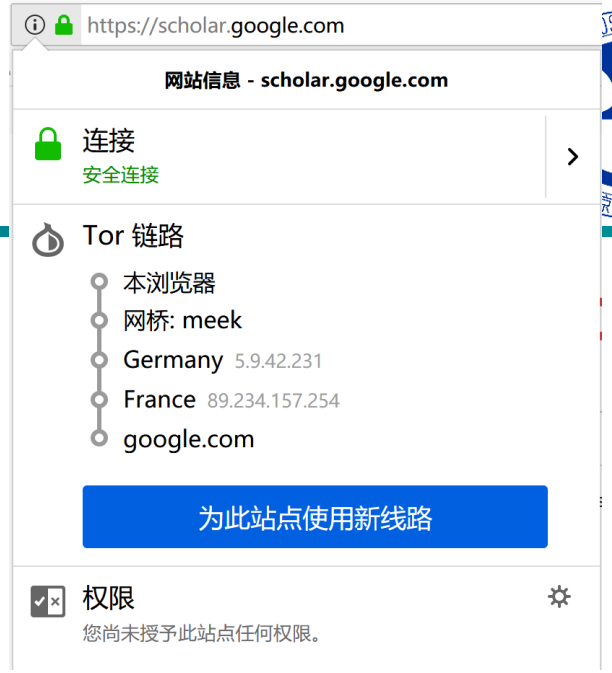


工作原理

基于影子身份实现服务与节点空间的映射，服务地址解析结果为影子身份的地址，基于多跳链路实现真实身份和影子身份的隔离





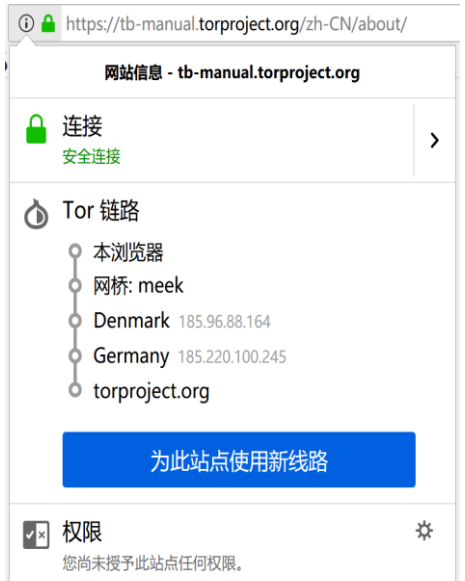


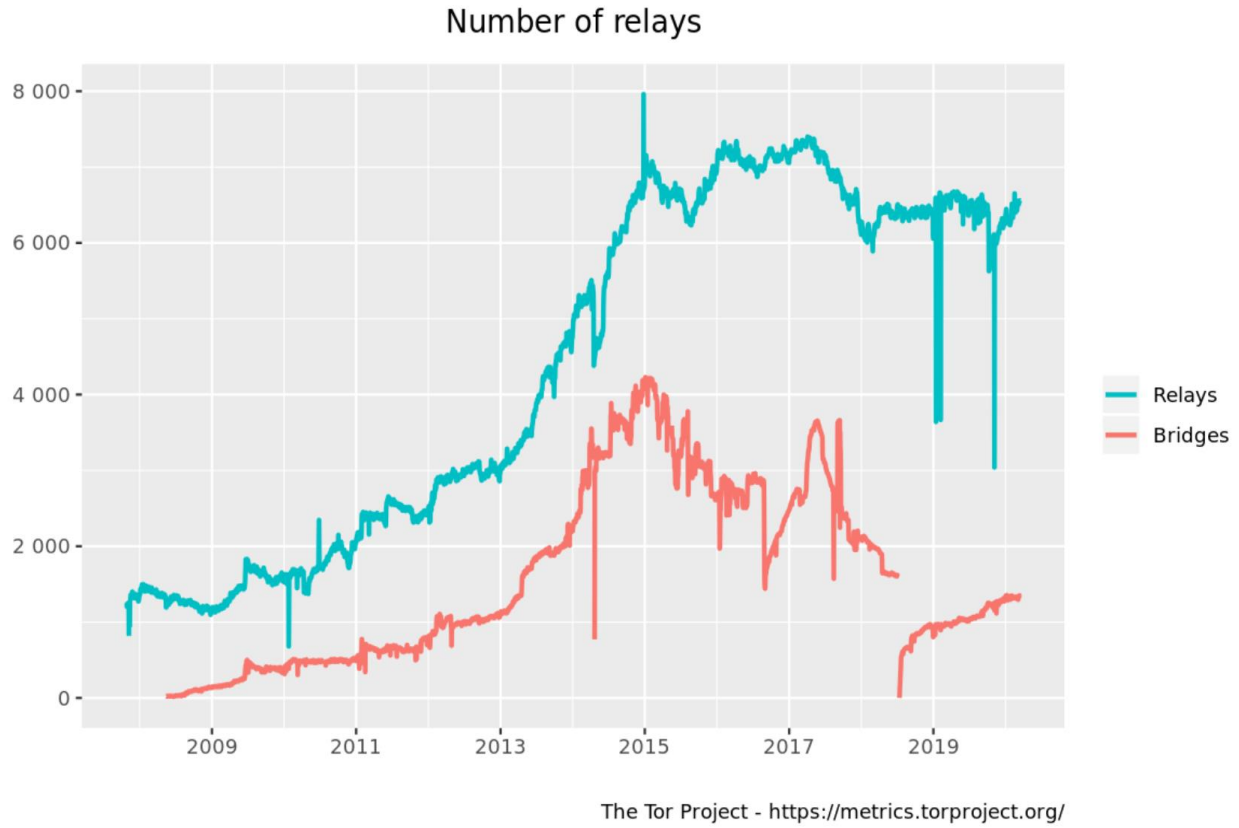
采用多跳隔离的方法关联用户/服务及其影子身份

部署与使用情况

- 2002年10月最初部署 Tor 网络时，其代码是在自由开放的软件许可下发布的。
- 截至2003年底，该网络有大约十二个志愿者节点，大多数在美国，还有一个在德国。
- SP 2006：450个中继节点

2007年，该组织开始开发接入 Tor 网络的桥接工具，以解决审查问题，例如绕过政府防火墙，以便其用户能访问开放网络。

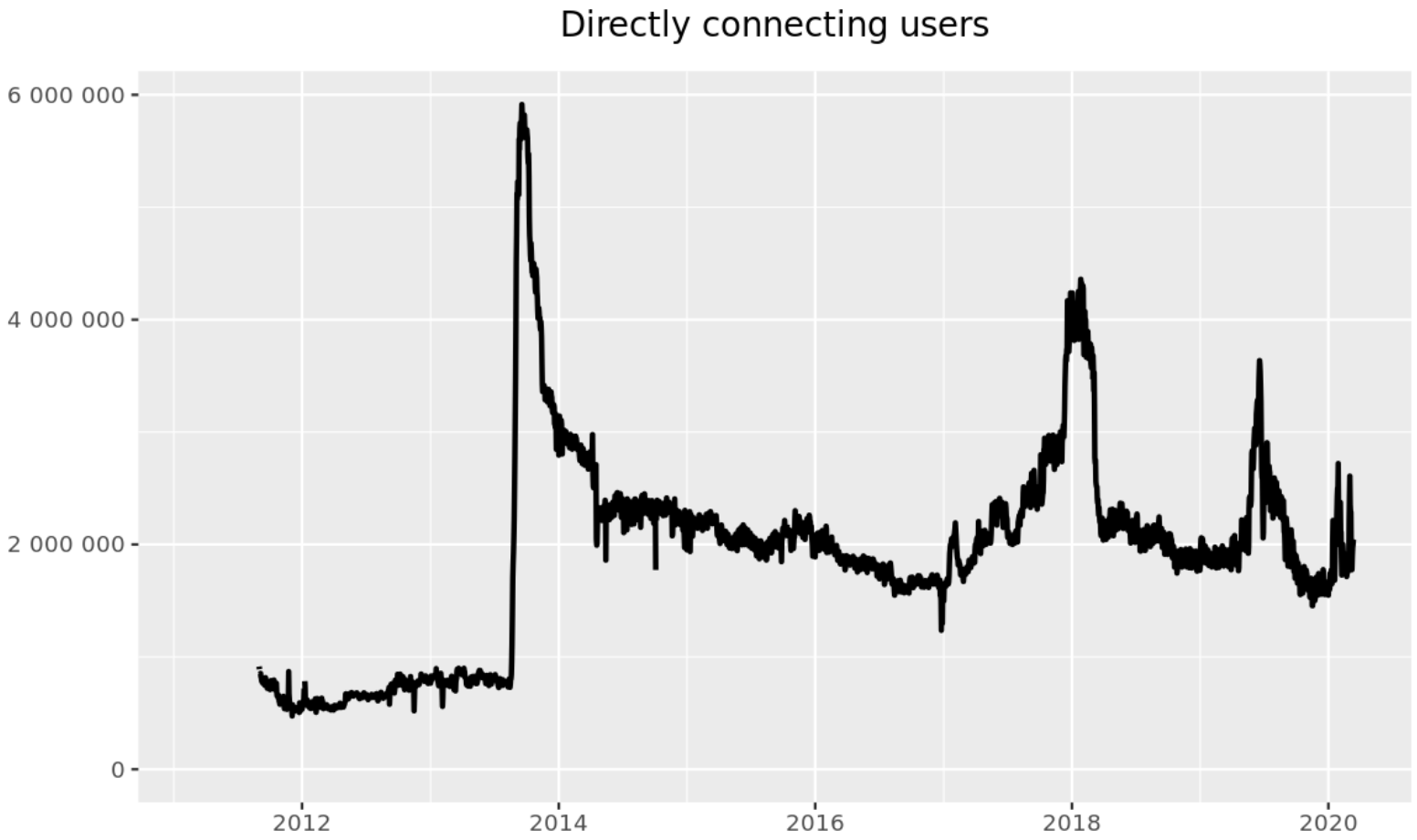




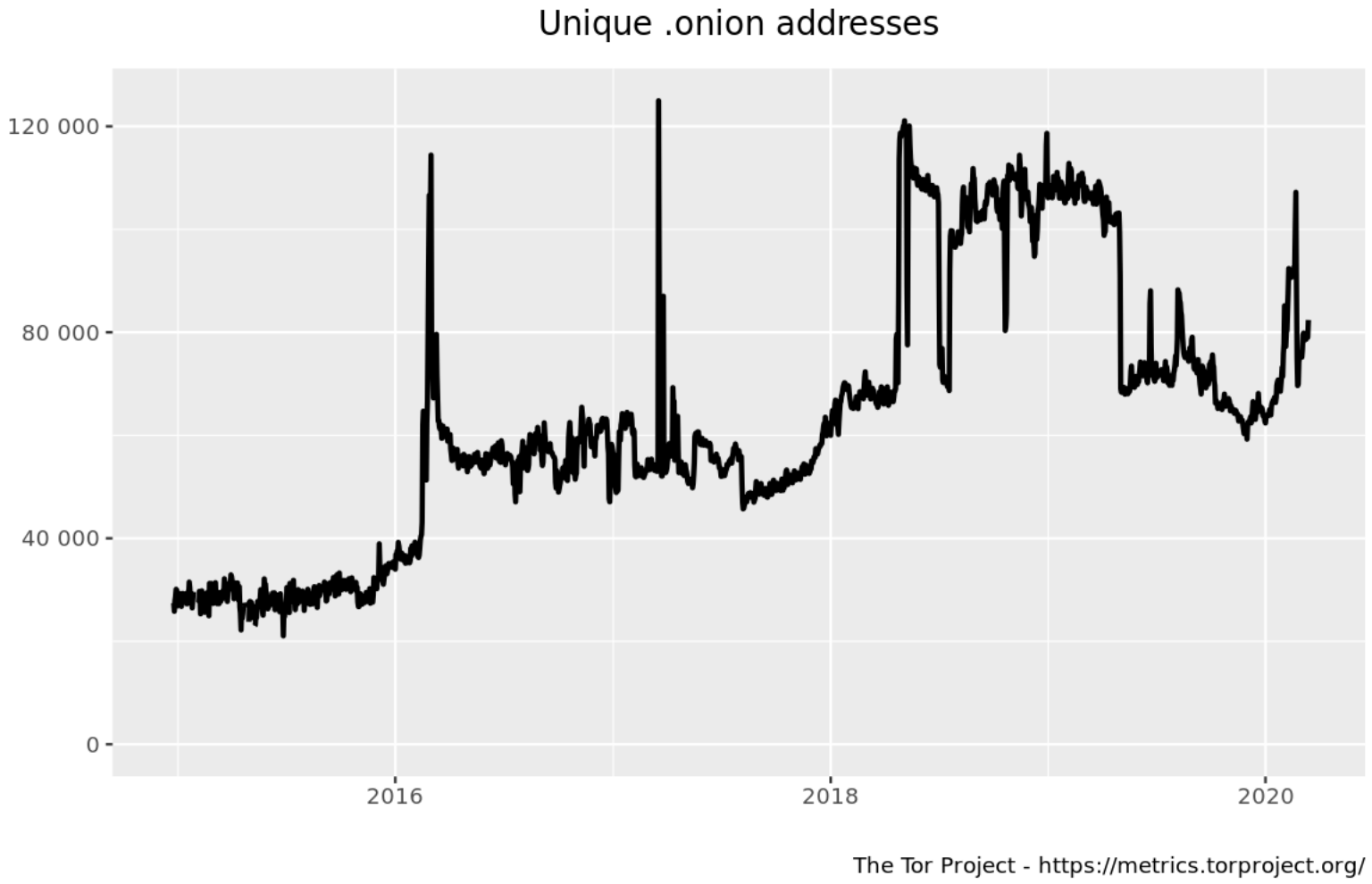
（2016.8.31），全球大概有7000个中继节点，2000个入口节点和1000个出口节点。

用户数量的变化

- 2013年，斯诺登(Snowden)事件

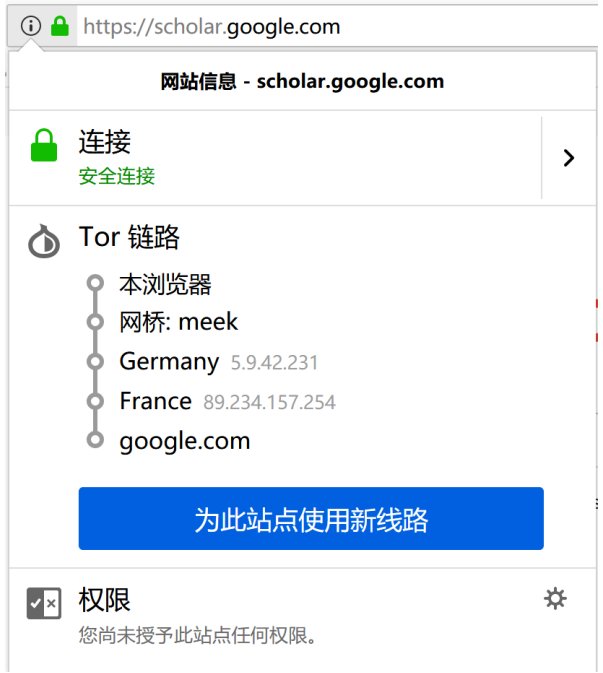


服务数量 - Tor Hidden Services (HS)



使用示例


bgbquu5ukeya2iuun7e7lv63ak2sttx3qcytdgf7mn2zouoruyceg6ad.onion
67%
网站信息 - bgbquu5ukeya2iuun7e7lv63ak2sttx3qcytdgf7mn2zouoruyceg6ad.onion
连接
连接不安全
Tor 链路
本浏览器
网桥: meek
Switzerland 80.253.92.71
United States 199.249.230.88
中继
中继
中继
bgbquu5ukeya2iuun7e7lv63ak2sttx3qcytdgf7mn2zouoruyceg6ad.onion
为此站点使用新线路
权限
您尚未授予此站点任何权限。


https://scholar.google.com
网站信息 - scholar.google.com
连接
安全连接
Tor 链路
本浏览器
网桥: meek
Germany 5.9.42.231
France 89.234.157.254
google.com
为此站点使用新线路
权限
您尚未授予此站点任何权限。

示例：茶马古道 – 中文暗网市场

1) Goldschlag, David M., Michael G. Reed, and **Paul F. Syverson**. "Hiding routing information." In *International workshop on information hiding*, pp. 137-150. Springer, Berlin, Heidelberg, 1996.

2) Reed, Michael G., **Paul F. Syverson**, and David M. Goldschlag. "Proxies for anonymous routing." In *Proceedings 12th Annual Computer Security Applications Conference*, pp. 95-104. IEEE, 1996.

3) Syverson, Paul F., David M. Goldschlag, and Michael G. Reed. "Anonymous connections and onion routing." In *Proceedings. 1997 IEEE Symposium on Security and Privacy (Cat. No. 97CB36097)*, pp. 44-54. IEEE, 1997.

4) Reed, Michael G., **Paul F. Syverson**, and David M. Goldschlag. "Anonymous connections and onion routing." *IEEE Journal on Selected areas in Communications* 16, no. 4 (1998): 482-494.

5) Goldschlag, David, Michael Reed, and **Paul Syverson**. "Onion routing." *Communications of the ACM* 42, no. 2 (1999): 39-41.

1) Syverson, Paul, Roger Dingledine, and Nick Mathewson. "Tor: The secondgeneration onion router." In *Usenix Security*, pp. 303-320. 2004.

2) Overlier, Lasse, and **Paul Syverson**. "Locating hidden servers." In *2006 IEEE Symposium on Security and Privacy (S&P'06)*, pp. 15-pp. IEEE, 2006.

3) Dingledine, Roger, Nick Mathewson, and **Paul Syverso**n. "Deploying low-latency anonymity: Design challenges and social factors." *IEEE Security & Privacy* 5, no. 5 (2007): 83-87.

1) Syverson, Paul. "A peel of onion." In *Proceedings of the 27th Annual Computer Security Applications Conference*, pp. 123-137. 2011.

2) Johnson, Aaron M., Paul Syverson, Roger Dingledine, and Nick Mathewson. "Trust-based anonymous communication: Adversary models and routing algorithms." In *Proceedings of the 18th ACM conference on Computer and communications security*, pp. 175-186. 2011.

3) Jansen, Rob, Paul Syverson, and Nicholas Hopper. "Throttling Tor bandwidth parasites." In *Presented as part of the 21st {USENIX} Security Symposium ({USENIX} Security 12)*, pp. 349-363. 2012.

4) Johnson, Aaron, Chris Wacek, Rob Jansen, Micah Sherr, and Paul Syverson. "Users get routed: Traffic correlation on Tor by realistic adversaries." In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pp. 337-348. 2013.

5) Jansen, Rob, John Geddes, Chris Wacek, Micah Sherr, and Paul Syverson. "Never Been {KIST}: Tor's Congestion Management Blossoms with Kernel-Informed Socket Transport." In *23rd {USENIX} Security Symposium ({USENIX} Security 14)*, pp. 127-142. 2014.

匿名性与带宽、延时的约束关系

- 通过混合网络进行匿名通信会带来计算和通信开销：
  - 用户消息被批量处理以创建一个匿名集(因此会延迟)，并且它们被填充或截断到标准长度以防止流量分析。

一个匿名通信协议只能实现以下三种属性中的两种:强匿名性(即强匿名性)、低带宽开销和低延迟开销

带宽开销、延迟开销和发送方匿名或接收方匿名与全局被动(网络级)对手之间的关系

对于给定数量的受损节点，我们在带宽和等待时间开销之间得出必要的约束，这些约束的违反使AC协议无法实现强匿名性。

TABLE I

Latency vs. bandwidth vs. strong anonymity of AC protocols, with the number of protocol-nodes K, number of clients N, and message-threshold T, expected latency ℓ' per node, dummy-message rate β.

Protocol	Latency	Bandwidth	Strong Anonymity
Tor [10]	$\theta(1)$	$\theta(1/\mathsf{N})$	impossible
Hornet [47]	$\theta(1)$	$\theta(1/\mathsf{N})$	impossible
Herd [48]	$\theta(1)$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Riposte [49]	$\theta(N)$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Vuvuzula [20]	$\theta(\mathsf{K})$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Riffle [21]	$\theta(\mathsf{K})$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Threshold mixes [14]	$\theta(T \cdot \mathsf{K})$	$\theta(1/\mathsf{N})$	impossible*
Loopix [24]	$\theta(\sqrt{K} \cdot \ell')$	$\theta(\beta)$	possible
DC-Net [15], [46]	$\theta(1)$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Dissent-AT [22]	$\theta(1)$	$\theta(\mathsf{N}/\mathsf{N})$	possible
DiceMix [16]	$\theta(1)$	$\theta(\mathsf{N}/\mathsf{N})$	possible

* if T in $o(poly(\eta))$

Das, Debajyoti, Sebastian Meiser, Esfandiar Mohammadi, and Aniket Kate. "Anonymity trilemma: Strong anonymity, low bandwidth overhead, low latency-choose two." In *2018 IEEE Symposium on Security and Privacy (SP)*, pp. 108-126. IEEE, 2018.

北京郵電大學
Beijing University of Posts and Telecommunications

感谢聆听！

匿名通信

- 基本概念
- 匿名通信的基本框架
- 技术方案

基本概念

- 匿名通信指采取一定的措施隐蔽通信流中的通信关系，使窃听者难以获取或推知通信双方的关系及内容。
- 匿名通信的目的就是隐蔽通信双方的身份或通信关系，保护网络用户的个人通信隐私，实现对用户在网络层的元数据的隐私保护。

匿名通信的基本框架

- 匿名通信的基本框架可以从三个方面加以阐述：匿名属性(anonymity property)、对手能力(adversary capability)和网络类型(network type)。

1.匿名属性
匿名属性包括不可辨识性(unidentifiability)和不可联系性(unlinkability)。
不可辨识性是指对手无法识别用户的身份和行为；
不可联系性是指对手无法通过观察系统将消息、行为和用户相关联。

匿名通信的基本框架 -对手能力(adversary capability)

2.对手能力
对手是意图降低、消除通信匿名的通信网络用户或用户的集合。匿名通信系统一般通过提出威胁模型(thread model)，来表明该系统能够抵抗的对手能力。
对手能力分为三个方面：可达能力(reachability)、攻击能力(attackability)和适应能力(adaptability)。

对手的可达能力分为全局（global）和本地（local）两种。具有全局能力的对手可以访问网络中所有的节点和链路，而具有本地能力的对手只能访问网络中部分的节点和链路。

攻击能力分为被动（passive）和主动（active）两种。攻击的目的是为了识别消息的发送者或接收者。
1）被动攻击一般由匿名通信网络的外部观测者发起，其主要行为为观测网络中传输的消息、网络中数据的流量，并通过对消息和流量的分析达到攻击的目的。
2）主动攻击一般由匿名通信网络的内部节点发起，其主要行为为通过其控制的部分通信节点修改通信消息、追溯通信行为、修改通信行为，来达到攻击的目的。

适应能力分为动态和静态两种。在匿名通信系统中，对手的适应能力一般是动态的，动态地跟踪网络的变化，实时地收集路径选择算法信息，实时地监控网络传输的消息和流量的变化。

匿名通信的基本框架 - 网络类型

匿名通信系统的网络类型由以下三个因素确定，分别为：路径拓扑（path topology）、路由机制（route scheme）和路径类型（path type）。

匿名通信系统的路径拓扑有两种，分别为：瀑布型（cascade）和自由型（free）。

1. 在瀑布型的网络中，发送者从匿名通信网络中选择固定的通信路径进行消息的传输；
2. 在自由型的网络中，发送者可以选择任意长度的通信路径进行消息的传输。
3. 一般意义上，自由型的网络拓扑比瀑布型的网络拓扑具有更强的匿名。

匿名通信系统的路由机制分为单播（unicast）、组播（multicast）、广播（broadcast）和任意播（anycast）。

目前基于系统效率和系统部署等实际问题的考虑，大多数实际部署的匿名通信系统的路由机制都是单播的机制。

匿名通信系统的路径类型分为简单（simple）和复杂（complex）。

简单的路径类型不允许出现路径的循环，中继的节点在整个路径中只能出现一次；

复杂的路径类型可以出现路径的循环，中继的节点在整个路径中可以出现多次。

技术方案

- 现阶段匿名通信的技术方案主要分为三类：
  - 基于Mix算法的匿名通信系统、
  - 基于洋葱路由算法的匿名通信系统
  - 基于泛洪算法的匿名通信系统。

基于Mix算法的匿名通信系统

- 基于Mix算法的匿名通信系统的核心思想是利用单个Mix节点或瀑布型的多个Mix节点实现匿名通信。
  - Mix节点是指网络中向其他节点提供匿名通信服务的节点，它接收用其公钥加密的数据，并对数据进行解密、批处理、重序、增加冗余字节等处理，然后将数据传输给下一个Mix或最终接收者。
- 基于Mix算法的匿名通信系统具有以下特点：

1)匿名通信系统网络中一部分节点为其他节点提供匿名通信服务；
2)发起者需要在发起匿名通信之前确定整个通信的传输路径，该路径在传输中不会改变；
3)发起者需要在发起匿名通信之前，得到整个传输路径中各个Mix节点的信息，包括地址、密钥信息等；
4)系统匿名较高，但通信传输的时延较高，一般不适合实时的数据通信。

洋葱路由

- 洋葱路由技术结合mix技术和agent代理机制，通过洋葱代理路由器，采用面向连接的传输技术，用源路由技术的思想对洋葱包所经过的路由节点进行层层加密封装。中间的洋葱路由器对所收到的洋葱包进行解密运算，得到下一跳洋葱路由的地址，剥去洋葱包的最外层，在包尾填充任意字符，使得洋葱包的大小不变，并将新的洋葱包传递给下一个洋葱路由器。

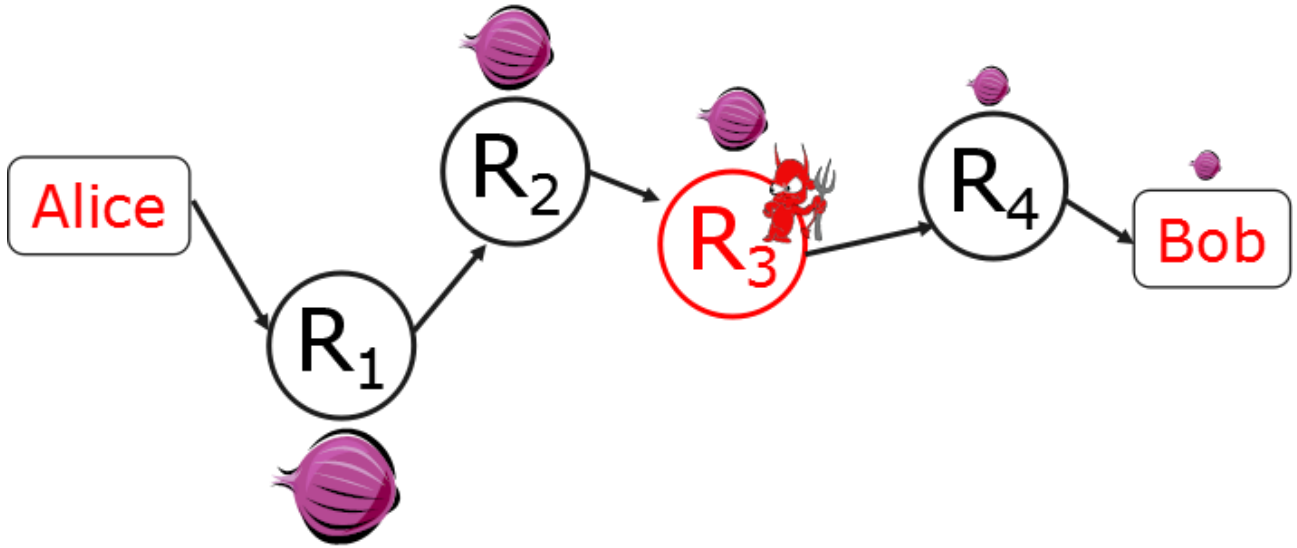
➢ 优势与不足：

- 洋葱路由技术为用户提供了双向、实时的匿名通信服务，但面向连接和严格的源路由方式导致存在效率低、扩展性差的不足

◆ 洋葱路由技术的特点

路由信息用路由节点的公钥加密，每个路由节点只能知道下一个路由节点的身份。

洋葱路由是由发送者控制路由路径的长度，并预先随机选择一条路由路径，而且某些路由节点是诚实的，某些可能是破坏者。



洋葱路由实现方式

洋葱路由

- 1)基于Onion Routing算法的匿名通信系统建立在TCP传输的基础上，节点之间通常通过SSL方式传输；
- 2)基于Onion Routing算法的匿名通信系统在路径建立时采用非对称密钥算法加密，在数据通信时采用对称密钥算法加密，以提高数据传输效率，降低时延；

基于泛洪算法的匿名通信系统

- 基于泛洪算法的匿名通信系统是近期匿名通信传输领域新的研究热点，主要基于flooding、epidemic等类洪泛算法实现匿名通信，目前仍处于实验室研究阶段，没有实际部署的成熟的匿名通信系统。
- 基于泛洪算法的匿名通信系统一般具有以下特点：
  - 1)发起者在发起匿名传输之前完全不清楚匿名传输的路径，也无需得到传输中间节点的任何信息；
  - 2)发起者的每一次匿名传输路径并不固定；
  - 3)匿名通信网络中的任何一个中间节点都不知道匿名通信的发起者和接收者。

基于泛洪算法的匿名通信系统主要面临的挑战是系统会产生大量的网络传输流量，对于网络带宽的需求较大；同时在目前的状态下，系统算法的稳定性和可靠性还不够。