Deploying Low-Latency Anonymity: Design Challenges and Social Factors

by Roger Dingledine (The Tor Project)
Nick Mathewson (The Tor Project)
Paul Syverson (US Naval Research Laboratory)

IEEE Security & Privacy, September/October 2007 (Vol. 5, No. 5),
pp. 83-87

Anonymous communication systems hide conversations against unwanted observations. Deploying an anonymous communications infrastructure presents surprises unlike those found in other types of systems. For example, given that users shouldn't need to trust each other or any part of the system, no single authority or organization should be able to observe complete traffic information for anyone's communication. This makes commercialization difficult and requires a rethinking of incentives for both users and infrastructure participants'in no small part because a user's security depends directly on the infrastructure's size and the number of other system users.

To address these and related issues, we designed Tor (the onion routing), a widely used low-latency, general-purpose anonymous communication infrastructure'an overlay network for anonymizing TCP streams over the real-world Internet. [1] Tor requires no special privileges or kernel modifications, needs little synchronization or coordination between nodes, and provides a reasonable trade-off between anonymity, usability, and efficiency. Since deployment in October 2003, the public Tor network has grown to about a thousand volunteer-operated nodes worldwide and traffic averaging more than 110 Mbytes per second from hundreds of thousands of concurrent users, ranging from ordinary citizens concerned about their privacy to law enforcement and government intelligence agencies looking to operate on the Internet without being noticed and corporations that don't want to reveal information to their competitors.

This article discusses how to use Tor, who uses it, how it works, why we designed it the way we did, and why that design makes it usable and stable.

I. Distributed trust and usability

The US Naval Research Laboratory and the Free Haven Project researched, developed, and deployed Tor, the third generation of deployed onion-routing designs, [1--3] under US Office of Naval Research (ONR) and DARPA funding to secure government communications. Two years after Tor's deployment in 2003, the Electronic Frontier Foundation (www.eff.org) funded Free Haven's continuing efforts for one year to help maintain ordinary citizens' civil liberties online. In 2006, the Tor Project incorporated as a nonprofit (www.torproject.org) and has received continued funding from the Omidyar Network, the US International Broadcasting Bureau, and other groups committed to fighting blocking and censorship on the Internet. This funding diversity fits Tor's overall philosophy---a wide variety of interests helps maintain the network's stability and

security.

Tor lets users connect to Internet sites without revealing their logical or physical locations to those sites or outside observers. Its locationhidden services also give publicly accessible hosts similar protection against being located. To connect to a remote server via Tor, the client software first gets a list of Tor nodes via a voting protocol from several central directory servers (to avoid dependence on or complete trust in any one server). It then incrementally creates a private pathway, or _circuit_, across the network via encrypted connections through authenticated Tor nodes whose public keys come from the directory servers. After choosing the nodes at random (subject to a preference for higher-performing nodes to allocate resources effectively), the client software negotiates a separate set of encryption keys for each hop along the circuit, beginning with a client-chosen preferred set of first nodes, called _entry guards_, to complicate profiling attacks by internal adversaries. [4]

The client software extends the circuit one node at a time, tunneling extensions through established portions of the circuit. Each node along the way knows only the immediately preceding and following nodes, so no individual Tor node knows the complete path that each fixed-sized data packet (or cell) will take. Thus, neither an eavesdropper nor a compromised node can see both the connection's source and destination. Later requests use new circuits to complicate long-term linkability between different actions by a single user.

Tor attempts to anonymize the transport layer, rather than the application layer. Thus, it can protect even authenticated communications via applications such as SSH. Moreover, Tor doesn't relay arbitrary IP packets; it can anonymize only TCP streams and DNS requests. Though limiting, this also means that Tor can rely on TCP's guaranteed in-order delivery, rather than rebuild such features for applications that use them. It also simplifies the cryptographic implementation. Some communication requires anonymity from a communication partner as well as from the network infrastructure. In such cases, if application-level protocols transmit identifying information, you can use additional scrubbing proxies, such as Privoxy for HTTP (www.privoxy.org).

In addition to providing security through Tor's distributed infrastructure and circuit design, usability is also a central goal. The Tor download comes with install wizards and GUIs for the major operating systems (GNU/Linux, Mac OS X, and Windows), and it also runs on various flavors of BSD and Unix. The basic instructions, documentation, FAQs, and so on are available in many languages. The Tor Vidalia GUI is designed to simplify server configuration (choosing exit policies, determining how much bandwidth to allocate to Tor, and so on). The Torbutton GUI offers Firefox users a one-click toggle to select whether or not browsing goes through Tor. A site administrator can easily configure the application to run at individual desktops, a site firewall, or a combination of the two.

The ideal Tor network would be practical, useful, and anonymous. When trade-offs arise among these properties, our research strategy has been to remain useful enough to attract many users and practical enough to support them. Only subject to these constraints do we try to

maximize anonymity. Tor's security and flexibility thus make it stand apart from other deployed traffic analysis resistance systems. Mix networks such as Mixminion [5] provide the highest degrees of practical anonymity, but that comes at the expense of highly variable delays that make such networks unsuitable for applications such as Web browsing. Commercial single-hop proxies (such as www.anonymizer.com) can provide good performance, but the single-point compromise can expose all users' traffic, and a single-point eavesdropper can perform traffic analysis on an entire network. Also, these proprietary implementations place any infrastructure that depends on these single-hop solutions at the mercy of its provider's financial health as well as network security.

Numerous other designs exist for distributed anonymous low-latency communication. Some have been deployed or even commercialized, [6,7] whereas others reside only on paper. [8,9] Each design offers something unique, but we feel that Tor has advantages that make it a superior choice for most users and applications. Unlike purely peer-to-peer (P2P) designs, for example, we neither limit ordinary users primarily to content and services available only within our network (as does www.i2p.net) nor require them to take responsibility for connections outside the network, unless they separately choose to run server nodes. [10] Nonetheless, because we support lowlatency interactive communications, end-to-end traffic-correlation attacks [11,12] are possible by an attacker who can observe both ends of a communication to correlate packet timing and volume, quickly linking the initiator to the destination.

Our defense rests in having a diverse enough set of nodes to let us distribute each transaction over several nodes in the network and prevent most real-world adversaries from being in the right places to attack users. This ``distributed trust'' approach means a wide variety of mutually distrustful users can safely operate and use the Tor network, thus providing sustainability and security. If most participating providers are reliable, Tor tolerates some hostile infiltration of the network. This distribution of trust is central to the Tor philosophy and pervades Tor at all levels:

- Onion routing has been open source since the mid-90s, thus letting mistrusting users inspect the code themselves.

- Tor is free software, and so anyone could take up its development from the current team.

- Anyone can use Tor without license or charge, which encourages a broad user base with diverse interests.

- Tor is designed to be usable, which also encourages a broad user base, and configurable, so that users can easily set up and run server nodes.

- Tor's infrastructure is run by volunteers scattered around the globe, which means it's neither dependent on any company's economic viability or business strategy nor completely under any one country's jurisdiction.

- The diversity of funding sources for ongoing development and

deployment helps ensure that the project isn't overly beholden to any one funder or to funders with any one primary purpose or even sources in any one jurisdiction.

All of these contribute to Tor's resilience and sustainability.

II. Social challenges

Many of the issues the Tor project needs to address extend beyond system design and technology development. In particular, the Tor project's image with respect to its users and the rest of the Internet impacts the security it can provide. With this issue in mind, we turn to the Tor user base and Tor's interaction with other services on the Internet.

Communicating security

Because it affects the possible anonymity set (that is, the number of other undistinguished communicants), usability contributes to anonymity systems' security.[13,14] Inversely, an unusable system attracts few users and thus can't provide much anonymity. To get the protection of a larger anonymity set, users should choose which anonymizing system to use based in part on how usable and secure others will find it. Thus we might supplement the adage ``usability is a security parameter'' [14] with a new one: ``perceived usability is a security parameter.'' [15]

Reputability and perceived social value

Another factor that impacts the network's security is its reputability---its perceived social value based on its current user base. If Alice is the only user who has ever downloaded the software, it might be socially accepted, but she's not getting much anonymity. Add a thousand activists, and she's anonymous, but everyone thinks she's an activist too. Add a thousand diverse citizens (cancer survivors, people concerned about identity theft, law enforcement agents, and so on) and now she's harder to profile.

Furthermore, the network's reputability affects its operator base: more people are willing to run a service if they believe it will be used by human rights workers, for example, than if they believe it will be used for disreputable ends. This effect is even stronger if node operators think they'll be associated with their users' ends.

So the more cancer survivors on Tor, the better the impact for the human rights activists. The more malicious hackers, the worse the effect on normal users. Thus, reputability is an anonymity issue for two reasons. First, sustainability is affected because a network constantly on the verge of being shut down cannot attract adequate nodes, which in turn affects performance and thus drives away users. Second, a disreputable network is more vulnerable to legal and political attacks because it will attract fewer defenders.

Reputability becomes even trickier with privacy networks because the good uses (such as publishing by journalists in dangerous countries, protecting road warriors from profiling and potential physical harm, tracking criminals, and protecting corporate research interests) are typically kept private, whereas network abuses or other problems tend to get wider publicity.

Abuse

For someone willing to be antisocial or even break the law, Tor is usually a poor choice for hiding bad behavior. For example, Tor nodes are publicly identified, unlike the million-node botnets that are now common on the Internet. Nonetheless, we've always expected that, alongside legitimate users, Tor would also attract troublemakers who exploit the network to abuse services on the Internet with vandalism, rude mail, and so on. To deal with such users, Tor is designed so that individual nodes can use exit policies to block access to specific IP/port ranges. This approach aims to make operators more willing to run Tor by letting them prevent their nodes from being used for abuse. For example, Tor nodes block SMTP (port 25) by default to avoid the issue of spam.

Yet, exit policies are useful but insufficient. If not all Tor nodes block exit to a given service, that service might try to block the entire Tor network instead. Although being blockable is important to being good netizens, we want to encourage services to allow anonymous access. Services shouldn't need to decide between blocking legitimate anonymous use and allowing unlimited abuse. Blocking IP addresses is a course-grained solution given that entire apartment buildings, campuses, and even countries sometimes share a single IP address. [16] Also, whether intended or not, such blocking supports the repression of free speech. In many locations where Internet access is censored or even punishable by imprisonment, Tor is a path both to the outside world and to others inside. Blocking posts from Tor makes the censoring authorities' jobs easier. This is a loss for both Tor and services, such as Wikipedia, which block Tor. We don't want to compete for (or divvy up) all the NATprotected entities of the world according to whether each contains a Tor (exit) node and thus gets blocked by Wikipedia. This is also unfortunate because relatively simple technical solutions exist that allow anonymous communication while curtailing abuse. [17]

For example, a service could prevent abuse and remove incentives for attempts to abuse by implementing various schemes for escrowing anonymous posts until editors reviewed them. As an extension, pseudonymous reputation tracking of posters through Tor could let users establish adequate reputations to post without escrow. [17,18]

We stress that, as far as we can tell, very few Tor uses are abusive. Few services have complained, and others are actively working to find ways other than banning to cope with the little abuse they have experienced. For example, the Freenode Internet Relay Chat (IRC) network had a problem with a coordinated group of abusers joining channels and subtly taking over the conversation. When Freenode labeled all users coming from Tor IP addresses as ``anonymous users,''

thus removing the ability to blend in, the abusers stopped using Tor. Simple technical mechanisms can remove the ability to abuse anonymously without undermining the ability to communicate anonymously. Tor is the largest and most diverse low-latency anonymity network available, but we're still in the early stages and several major questions remain.

First, will our volunteer-based approach to sustainability continue to work as well in the long term as it has the first several years? In addition to node operation, volunteers are increasingly taking on Tor research, deployment, maintenance, and development. Tasks include package maintenance for various OSs, document translation, GUI design and implementation, development of live CDs, and specification of new design changes.

What's more, Tor is only one of many components that preserve privacy online. For circumstances in which it's desirable to keep identifying information out of application traffic, someone must build more and better protocol-aware proxies that ordinary people can use. We also need to maintain a reputation for social good and to learn to coexist with the variety of Internet services and their established authentication mechanisms. We can't just keep escalating the blacklist standoff forever.

Finally, the current Tor architecture hardly scales to handle current user demand. We must deploy designs and incentives to further encourage clients to also relay traffic without thereby trading away too much anonymity or other properties. These open questions are challenging, but choosing not to solve them means leaving most users to a less secure network or without any anonymizing network at all.

Acknowledgments:
Thanks to Matt Edman for many helpful comments on a draft of this article.

References

1. R. Dingledine, N. Mathewson, and P. Syverson, ``Tor: The Second-Generation Onion Router,'' Proc. 13th Usenix Security Symp., Usenix Assoc., 2004, pp. 303--319; http://tor.eff.org/tor-design.pdf.

2. D.M. Goldschlag, M.G. Reed, and P.F. Syverson, ``Hiding Routing Information,'' Information Hiding---1st Int'l Workshop, R. Anderson, ed., LNCS 1174, Springer-Verlag, 1996, pp. 137--150.

3. M.G. Reed, P.F. Syverson, and D.M. Goldschlag, ``Anonymous Connections and Onion Routing,'' IEEE J. Selected Areas in Comm., vol. 16, no. 4, 1998, pp. 482--494.

4. L. Øverlier and P. Syverson, ``Locating Hidden Servers,'' Proc. 2006 IEEE Symp. Security and Privacy, IEEE CS Press, 2006, pp. 100--114.

5. G. Danezis, R. Dingledine, and N. Mathewson, ``Mixminion: Design of a Type III Anonymous Remailer Protocol,'' Proc. 2003 IEEE Symp. Security and Privacy, IEEE CS Press, 2003, pp. 2--15.

6. O. Berthold, H. Federrath, and S. Köpsell, ``Web MIXes: A System for Anonymous and Unobservable Internet Access,'' Designing Privacy Enhancing Technologies: Workshop on Design Issue in Anonymity and Unobservability, H. Federrath, ed., LNCS 2009, Springer-Verlag, 2000, pp. 30--45.

7. A. Back, I. Goldberg, and A. Shostack, ``Freedom Systems 2.1 Security Issues and Analysis,'' white paper, Zero Knowledge Systems, May 2001.

8. M.J. Freedman and R. Morris, ``Tarzan: A Peer-to-Peer Anonymizing Network Layer,'' Proc. 9th ACM Conf. Computer and Comm. Security (CCS 02), ACM Press, 2002, pp. 193--206.

9. M. Rennhard and B. Plattner, ``Practical Anonymity for the Masses with Morphmix,'' Financial Cryptography, A. Juels, ed., Springer-Verlag, 2004.

10. M.K. Reiter and A.D. Rubin, ``Crowds: Anonymity for Web Transactions,'' ACM Trans. Information and System Security, vol. 1, no. 1, 1998, pp. 66--92.

11. G. Danezis, ``The Traffic Analysis of Continuous-Time Mixes,'' Proc. Privacy-Enhancing Technologies (PET 2004), D. Martin and A. Serjantov, eds., LNCS 3424, 2004; www.cl.cam.ac.uk/users/gd216/cmm2.pdf.

12. A. Serjantov and P. Sewell, ``Passive Attack Analysis for Connection- Based Anonymity Systems,'' Proc. 8th European Symp. Research in Computer Security (ESORICS), LNCS 2808, Springer-Verlag, 2003, pp. 116--131.

13. A. Acquisti, R. Dingledine, and P. Syverson, ``On the Economics of Anonymity,'' Financial Cryptography, R.N. Wright, ed., LNCS 2742, Springer-Verlag, 2003, pp. 84--102.

14. A. Back, U. Möller, and A. Stiglic, ``Traffic Analysis Attacks and Trade-Offs in Anonymity Providing Systems,'' Proc. Information Hiding (IH 2001), I.S. Moskowitz, ed., LNCS 2137, Springer-Verlag, 2001, pp. 245--257.

15. R. Dingledine and N. Mathewson, ``Anonymity Loves Company: Usability and the Network Effect,'' Designing Security Systems That People Can Use, O'Reilly Media, 2005, pp. 547--559.

16. G. Goodell and P. Syverson, ``The Right Place at the Right Time: Examining the Use of Network Location in Authentication and Abuse Prevention,'' Comm. ACM, vol. 50, no. 5, 2007, pp. 113--117.

17. J. Holt, ``Nym: Practical Pseudonymity for Anonymous Networks,'' white paper, 2005; www.lunkwill.org/src/nym/.

18. P.C. Johnson et al., ``Nymble: Anonymous IP-Address Blocking,'' Proc. Privacy-Enhancing Technologies (PET 07), Springer-Verlag, 2007.

Roger Dingledine is the president and cofounder of the Tor

Project. His research interests include security and scalable secure systems, anonymity and pseudonymity, cryptography, usability, the economics of privacy, and free software advocacy. Dingledine has an MEng in electrical engineering and computer science from MIT. He organizes academic conferences on anonymity, speaks at many industry and hacker events, and does tutorials on anonymity for national and foreign law enforcement. Contact him at arma@torproject.org.

Nick Mathewson is the chief architect at the Tor Project. His research interests include traffic analysis and communications anonymity. Mathewson has an M.Eng in computer science from MIT. He wrote the Mixminion next-generation anonymous remailer and helps maintain the Free Haven Anonymity Bibliography (www.freehaven.net/anonbib/). Contact him at nickm@torproject.org.

Paul Syverson is a mathematician at the US Naval Research Laboratory. His research interests include theory, design, and analysis of systems and protocols for anonymity, security, and privacy. Syverson has a PhD in philosophy and MAs in mathematics and philosophy from Indiana University. He is author of Logic, Convention, and Common Knowledge: A Conventionalist Account of Logic (CSLI Publications, 2003). Contact him at onion-info@itd.nrl.navy.mil.