Why I'm Not an Entropist

Paul Syverson*

Naval Research Laboratory
syverson@itd.nrl.navy.mil

Abstract. What does it mean to be anonymous in network communications? Our central thesis is that both the theoretical literature and the deployed systems have gotten the answer essentially wrong. The answers have been wrong because they apply the wrong metric to the wrong adversary model. I indicate problems in the established adversary models and metrics for anonymity as well as implications for the design and analysis of anonymous communication systems.

1 Introduction

Anonymous communication is the quintessence of brief encounter. For anyone to whom it is anonymous, such communication cannot be linked to an individual, for example to its source or destination. But it also cannot be linked to any other instances of communication. If it could, it would be part of a communication profile, hence pseudonymous communication rather than anonymous. Anonymity thus guarantees that an encounter is brief rather than leaving this to chance.

Anonymity is also an area of security that is much younger than, for example, confidentiality. Mechanisms and networks for anonymous communication have been designed for only about thirty years [6] and deployed for not quite half that time [21,35,19]. We do not have nearly as much experience with deployed systems on which to build theory, definitions, and models as many other areas of security, for example confidentiality or authentication. Though the field has developed and adapted, the basic conception of what it means to be anonymous has persisted throughout its history.

That conception is of anonymity as about indistinguishability within a set. The idea is thus that there is a set of possible senders, and the adversary cannot adequately distinguish amongst the members of that set. (For brevity and convenience I focus discussion on sender anonymity of a message.) I call this *the entropist conception* of anonymous communication. I am motivated in choice of terminology here by general Rényi entropy rather than Shannon entropy specifically (H_2 in the ordering of Rényi entropies). Thus I am including everything from simple cardinality of the set of possible senders (the log of which is Hartley entropy, H_0) to the other extreme of min entropy (H_∞). However, I am more

* A version of this paper was originally presented to the 17*th* Security Protocols Workshop, Cambridge UK, April 2009, which had as its theme "brief encounters". This is a substantially revised and expanded version of the originally presented paper.

B. Christianson et al. (Eds.): Security Protocols 2009, LNCS 7028, pp. 213–230, 2013.

focused on the conception of anonymity as indistinguishability within a set (possibly according to some probability distribution) than on any particular mathematical characterization of it. I will argue that this basic conception—a conception that underlies virtually all extent work on anonymous communication—is not appropriate for the fundamental role it has been given.

Before getting into details, I set out some general principles for security metrics and system design that I believe should be uncontroversial. It is on the basis of these principles that I will make my case.

- To be useful, security metrics should reflect the difficulty an adversary has in overcoming them. (So, for example, a metric for a security property should not depend on variables whose values do no significantly impact whether the adversary succeeds in attacks on that property.)
- To be meaningful, security metrics should not depend on the values of variables for which we cannot make adequate relevant determinations or predictions. (It might be sufficient to determine or predict bounds, distributions, etc. for those values rather than the exact values themselves, provided they are relevant. For example, to observe that a value will always be finite will usually not reflect a bound that is relevant to a practical security metric.)
- Though security through obscurity is rarely a good system design strategy, if your design provides your adversary with an explicit target that he has available resources to overcome, you are doomed to fail.

It is crucial to understand that our fundamental complaint against entropism is *not* that entropy fails to ever properly say anything about the uncertainty we may have about senders or receivers of messages. The problem is that entropism fails to tell us much about the adversary's amount of knowledge (or lack of knowledge) that is useful in saying how secure our practical anonymity systems are, or how secure a given use of one of them is. Starting from an entropist conception of anonymity has led to a focus on system properties that are not the most important to system security, has led to system assumptions that are not reasonable in practice, has led to adversary models that are not reasonable in practice, and has led to both theory and system design for anonymous communication that fail in significant ways. The remainder of the paper investigates these points in light of the above principles. We concentrate herein on problems with the existing approach, but we will at least briefly consider potential alternatives to replace it with.

2 Wait. I Thought You Said You're *Not* an Entropist.

There are many circumstances where the entropist conception of anonymity makes perfect sense, for example, in an election where there are registered voters casting ballots. An entropist approach can protect the anonymity of voters by confounding of associating voters with ballots and provides a good metric for that protection. (This assumes that appropriate other protections ensure that each registration represents exactly one voter, that each voter casts at most one

ballot per election, and assumes that enough eligible voters vote, that enough voters cast ballots for each of the options—if that is of concern, etc.)

As a more exotic example, in the 1970s the United States and the Soviet Union sought to limit the number of nuclear missiles they both had through cooperation and inspection while maintaining security guarantees. A large part of one of the major initiatives involved shuttling Minuteman missiles about in a field of silos to preclude successful first-strike attacks directly on all available missiles. The plan also included techniques for communicating to the adversary an authenticated, integrity-protected report from each silo. The report indicated whether a missile was in the silo or not, but without the adversary being able to determine which silo the report was from (except by silo-unique but random identifier). In a field of 1000 missile silos, the adversary could be sure that exactly 100 would be occupied but would not know which ones were occupied. Note that this anonymity system used "dummy packets" in the form of huge "transportainer" trucks continually shuttling either actual missiles or dummy loads between silos. Talk about system overhead! See [40] for more details about the plan.

As a less exotic example, suppose an intimidating colleague or perhaps the boss is repeatedly but obliviously doing something that is inappropriate and bothers several coworkers: leaving the food area or bathroom a mess, telling jokes or making comments that some find offensive, etc. Nobody wants to be the one to complain to the offender. Also they don't want to embarrass or perhaps encourage ire by giving the impression that this was a topic of full group discussion and decision making, whether or not that is true. They just want it to stop. Thus they wouldn't want to use a collaborative anonymity mechanism to send this message unless it was also being used to send others. Put differently, they want the recipient to know that one of them sent the message, rather than that all of them sent it, even though only one copy was delivered.

Finally, a similar example that comes closer to our intended topic below. Suppose it is important to release some information to the press or the public without attribution. The information is posted to a public site. It is known that only a relatively small number of people had prior access to the information, so it was one of them, but we don't know which one. (Let us assume and ignore that the relevant set of people is known, and that there is no doubt that they are the only ones who had the information. Also assume that the authenticity of the information is evident or confirmed once revealed.) It could be government information (good or bad) of important public interest where overt indication of which knowledgeable person released the information would be a distraction from the public good of the release. Alternatively, perhaps the release is intentional and important but cannot be overtly approved for some reason, and if the "leaker" were known, s/he would be punished, officially or otherwise.

Entropism is about using entropy as the meaning, the criterion for anonymity or how anonymous something is. That has drawbacks because, as an average, it necessarily does not capture everything important. To the extent that entropy is the measure of anonymity it drives system design to maximize entropy, which may not be the same thing as maximizing anonymity in ways that matter. It is

just one number, or a few numbers, such as the Shannon entropy and the min entropy. Any of these might be fine as a measure of anonymity in an appropriate context. None is fine as *the* measure of anonymity. As *the* measure it effectively becomes the definition for a relatively complex concept which is no better served by limiting to any single definition than is the overall field security by limiting to a single definition of 'security'.

In all of the just given cases, there is a set of possible sources (or locations) and the relevant security issue is the uncertainty within that set. For some of the cases, the exact set may not be known, but reasonable bounds can be given for both set size and uncertainty. Uncertainty on a known or approximately known set is clearly the significant concern in these cases, but what are examples where entropy is not appropriate? We will argue that entropism is the wrong approach to design and evaluation of the most well known and widely used systems for protecting the anonymity of communication.

3 Protecting Anonymity

Tor is a very widely used and deployed onion-routing network, which means that it gets its protection from creating cryptographic circuits along routes that an adversary is unlikely to observe or control. As in previous onion-routing systems, in the currently deployed Tor design this is achieved by choosing unpredictable routes through the network. This could, however, also derive in part from the inherent difficulty an adversary has attacking a particular part of the network [25,26] or from other sources. We will return to this below. Also like other onion-routing systems, Tor passes traffic bidirectionally along those circuits with minimal latency [44]. It is common in the literature to characterize the anonymity protection Tor provides in entropist terms. For example, "[t]he anonymity provided by Tor relies on the size of the anonymity set. Currently, there are around 1500 ORs, and an estimated quarter million Tor users." [43].

Assume I am a user of the network who picks his entry node the same way the vast majority of users do and then visits a hostile website. On the entropist conception, my anonymity is the same as any other user of the network. With about a quarter million active users that gives me a (Shannon) entropy of about 18 with respect to being identified as going to this website. But the number of users is only an estimate, and the system is designed so that nobody should be able to actually see connections coming from all of them. Further, talking about the number of users ignores how attacks actually happen.

An adversary observing an entry node and an exit node of the Tor network through which I am, e.g., browsing the web can trivially link the two ends of the connection and correlate source to destination. This has been an acknowledged feature of the design since its inception [14]. By the same timing correlation, an adversary that controls or observes a destination website and the entry node will be able to identify the source address of the user accessing the website. Against these attacks the number of other users is all but irrelevant.

In this case, entropy fails our second criterion for a security metric. We cannot determine the number of users on the network in general or the number using circuits with the same entry and exit nodes at the time of the attack. We can estimate these numbers based on network measurements such as given above. That would, however, only tell us the average number of users during various periods in the past. What should matter for my (entropist) anonymity with respect to the adversary is my distinguishability from other users during the attack. If we cannot know when the attack occurs, could we give at least a lower bound on the anonymity set size during any reasonable candidate attack period? The answer is no. This means that we cannot use set size or indistinguishability within a set as a metric of my security when using the system. Could it be because nobody knows the anonymity set size so that my protection is absolute?

No. Entropy also fails our first criterion for a useful security metric: it should reflect the difficulty an adversary has in overcoming security and should not depend on variables whose values do not significantly impact the adversary's success. We could imagine rare cases where two Tor circuits through the same entry and exit nodes might be hard to distinguish by at least a completely passive adversary, but it is well known and accepted that whether you are the only user of those nodes during the attack or there are a hundred others, linking your source and destination is trivial. So the anonymity set size cannot be determined, and the difficulty of the most salient attacks are not affected by what it might be.

Perhaps we are discussing uncertainty on the wrong set. Perhaps the number of users has only an indirect effect on the security of Tor traffic. We have repeatedly observed that the primary threat we are discussing is the end-to-end timing correlation. The relevant set is thus the number of locations at which traffic might enter and exit the network.

In the scenario in which I go to a hostile website, what will matter is the nodes through which I can enter the system. If I make that choice the same way as everybody else, then I currently have on the order of three hundred choices.[1] So an entropy of a little over eight. But that assumes that the probability of entry node choice is uniform. This is not actually true. It is weighted by node bandwidth and other factors that we will not go into. We could still calculate the entropy of a typical connection entering the network. But even if we ignore that and assume the choice of entry nodes was uniform we must ignore something else as well—how likely a node is to be compromised. Suppose that, instead of choosing uniformly (or possibly weighted to help maintain performance), I always choose my entry nodes to be just one of a handful of nodes. If the nodes in that handful are all run by highly trusted and technically competent friends who would never betray me, my anonymity is clearly better than picking uniformly from the much larger set: the hostile website will never identify me through the

[1] For those who know about entry guards, we are describing the choice of entry guards not the subsequent choice of route once guards are chosen. For those who do not know about entry guards, please do not be distracted by this footnote.

cooperation of a hostile entry node.[2] My anonymity protection has far more to do with the security of that first node I choose in my path than with the number of other users or even other possible first nodes. Again, this is a variable that does not affect the difficulty of attacking my anonymity.

Using entropy as a security metric for anonymity conflates the adversary's uncertainty with the effort needed to reduce that uncertainty. This conflation is fine as long as the effort needed to reduce uncertainty is more or less uniform in the degree of uncertainty (as would be the case for, e.g., key length for a well designed block cipher). When it is not, failing to focus on anonymity protection distorts both security analysis and secure system design.[3]

Even within the entropist conception something akin to the distinction between entropistic anonymity and anonymity protection is recognized. In [37], Pfitzmann and Hansen distinguish between anonymity and strength of anonymity: "Robustness of anonymity characterizes how stable the quantity of anonymity is against changes in the particular setting, e.g., a stronger attacker or different probability distributions. We might use quality of anonymity as a term comprising both quantity and robustness of anonymity. To keep this text as simple as possible, we will mainly discuss the quantity of anonymity in the following, using the wording 'strength of anonymity'." But they still make entropy the primary measure of how well anonymity is protected. "All other things being equal, global anonymity is the stronger, the larger the respective anonymity set is and the more evenly distributed the sending or receiving, respectively, of the subjects within that set is." They do note how individuals may have weak anonymity regardless of the average uncertainty across the anonymity set. "Even if global anonymity is strong, one (or a few) individual subjects might be quite likely, so their anonymity is weak. W.r.t. these 'likely suspects', nothing is changed if the anonymity set is made larger and sending and receiving of the other subjects are, e.g., distributed evenly. That way, arbitrarily strong global anonymity can be achieved without doing anything for the 'likely suspects' [7]." This distinction in [7], however, is intended to cope with the influence of statistical outliers on average uncertainty within a set versus uncertainty concerning an individual and thus to suggest using quantiles as a metric for those cases. It does not at all question the basic idea that anonymity is based on a level of uncertainty from within a known set.

To reiterate, the problem with the entropist conception is not that entropy entirely fails to reflect uncertainty (once the situation is properly described). The problem is that focusing on the size of a known set of users and their distinguishability within it skews system design as well as what gets analyzed in a way that obscures rather than enhances our understanding of the anonymity

[2] Whether or not these nodes are known to be affiliated with me now becomes relevant, which depends on the probability that the second node in my path is compromised (and ignoring links as possible places of compromise) [34]. We will touch briefly on such issues below.

[3] For a cryptographic analogue see the discussion of locally unpredictable delaying functions in [20] wherein most of the bits of a function might be easy to compute, but a few of them require much more effort.

protection provided by our systems. Wanting to communicate anonymously and so making entropy the focus of your anonymity design is like wanting to be married and so making expected-number-of-spouses the focus of your matrimony design. If that is indeed you primary focus, you are likely to end up with a strategy of, e.g., proposing indiscriminately to as many people as possible.[4]

If probability of node compromise could be assumed to be roughly uniform (and ignoring links), then network size would be a primary determiner of anonymity protection. Similarly, if we could produce bigger, more uniform looking (to the adversary) sets of senders and recipients, that might actually be useful. But that is not a realistic view of how large, widely used anonymity networks work. They are comprised of fairly dynamic and diverse collections of users communicating over nodes that are diversely trusted by diverse parties and that are diversely configured versions of diverse platforms. And these nodes are connected over diversely trusted links (based on ASes, ISPs, geography, etc.). Unlike designing a closed secure system, there is no point in even discussing trying to make the degree of security of all of the different parts of the network roughly comparable. We will touch on more centrally operated systems below. We will also see that the number of nodes in an open system does play a role in the protection a system offers, but only in combination with the resistance those various nodes have against adversaries.

In security analysis, most adversaries are worst-case, possibly subject to some constraints on ability. But they can attack any message anywhere. 'Attack' might just mean observe, but the adversary can be in the worst possible place. If there is a subset of nodes, or messages that he is able to attack, it is assumed that these can be the optimally most effective nodes or messages. We will next consider what an adversary can do and where.

4 What Is an Adversary?

"*Keep your friends close, and your enemies closer.*"

Sun-tzu — Chinese general and military strategist (c. 400 BCE)

Computer security, including anonymity, is fundamentally tied up with the idea of an adversary. Indeed the primary difference between security and other types of computer assurance is the assumption in security that there is a potential for intentional misuse of some sort as opposed to simply accidental errors or defects. An old security maxim puts it this way: a system without an adversary model cannot be insecure, just surprising.

Research into design and analysis of anonymous communication began in the cryptologic community where vetting and deploying cryptosystems and protocols takes a long time and where a practical break in a deployed system can have serious and widespread consequences. A nice example of this is the evolution from initially discovered weaknesses to practical attacks on MD5. Some vulnerabilities are only of theoretical importance, however. Though they may

[4] Thanks to Nick Mathewson for this analogy.

motivate further research, they will never themselves be implemented in a significant practical attack even if systems are never modified in response to them. They will never be the lowest hanging fruit. Contrast this with the Dolev-Yao adversary model [17], an attack against which typically implies dangerous real-world attacks against any implementation of a protocol that is vulnerable to it. Anonymous communications research started with the crypto approach of looking at how hard it is to break the design of an isolated component that might be crucial. The community has still not fully recognized the very limited usefulness such an approach would have for designing and analyzing practical widescale systems for anonymous communication.

Mix networks get their security from the mixing done by their component mixes, and may or may not use route unpredictability to enhance security. Onion routing networks primarily get their security from choosing routes that are difficult for the adversary to observe, which for designs deployed to date has meant choosing unpredictable routes through a network. And onion routers typically employ no mixing at all. This gets at the essence of the two even if it is a bit too quick on both sides.[5] Mixes are also usually intended to resist an adversary that can observe all traffic everywhere and, in some threat models, to actively change traffic. Onion routing assumes that an adversary who observes both ends of a communication path will completely break the anonymity of its traffic. Thus, onion routing networks are designed to resist a local adversary, one that can only see a subset of the network and the traffic on it.

Given the fundamental differences in the mechanisms they employ, the adversaries they are intended to resist, and their basic designs (not to mention typical applications) it might seem impossible or at least astonishing that anyone who works in this area would ever confuse the two. Yet for years it has been common for publications by even top researchers in anonymous communication to refer to onion routing networks as mixnets or vice versa. All deployed onion routing networks do use some form of layered encryption on traffic they carry, encryption that is gradually removed as it passes through the network. And this is also true of decryption mixnets (re-encryption mixes work differently). Thus there is a clear similarity between the two in at least this respect. Still, given the differences, it would be surprising if this were enough to confuse an expert. But, if you start from an entropist conception of anonymity the confusion becomes less surprising: if you start from an entropist conception of anonymity, all anonymity designs are trying to make a given set of users (or user communications) less distinguishable. If one motivates design by starting with such a set and seeing how well the system obscures identification of its elements, the security contributions of an onion routing approach are harder to see. Any distinction between onion

[5] Other typical and highly salient distinctions include that all existing onion routing network designs are for carrying bidirectional low-latency traffic over cryptographic circuits while public mixnets are designed for carrying unidirectional high-latency traffic in connectionless messages. (An exception is the Web MIXes design [3] as deployed in JonDonym [27], which creates bidirectional circuits through a mix cascade to carry public web traffic.)

routing networks and mixnets, if recognized at all, is then likely to be couched only in terms of differences in intended application or engineering tradeoffs of security (in the entropist model) versus performance. Even we designers of onion routing systems have been occasionally guilty of falling into this idiom.

But does using the entropist conception reveal potential low-hanging fruit for current or future real systems? To answer this we should consider adversaries that can conduct practical attacks. For brevity's sake, I describe only the one that matters most in examining if entropism is the best approach for widely-used systems like Tor [44], Mixmaster [30], and the Anonymizer [2], viz: **The Man**.[6]

The Man owns big chunks of the anonymity infrastructure, either because he simply set them up himself, or because they are not hardened against takeover. He can also get access to ISPs, backbones, and websites, will know ancillary things, and, if targeting you, will have you specifically under physical surveillance. Think organized crime, state level actors (intelligence, secret police), etc. The Man subsumes the other adversaries we might consider.

In the literature, a standard anonymity adversary is the global-passive adversary (GPA), who controls no nodes in the anonymity network but can observe all traffic on all links to and from that network as well as between nodes of the network. This adversary can observe all sending and receiving behavior by all principals interacting with the network. It thus fits nicely in the entropist model of anonymity and facilitates formal analysis therein. While nice for producing both positive and negative provable results, the GPA is too strong to be realistic because it can observe absolutely every link everywhere regardless of network size. Good security counsels a conservative view, assuming an attacker that is stronger than anything one would encounter in practice. The GPA might thus be considered an overstatement of The Man—except that the GPA is also much too weak; it cannot even briefly delay a packet passing over a single link under its observation. As has long been recognized, the GPA is both too strong and too weak for low-latency distributed systems like onion routing [38,42,14].

All low-latency systems as currently designed and deployed are essentially broken against The Man, but often much weaker adversaries are adequate. Single proxies are vulnerable to timing attacks by much weaker adversaries, for example, a single well-placed network observer, and of course communication is vulnerable to the proxy itself if it is corrupt or compromised. Realtime web cascades, such as JonDonym [27] are generally to be comprised of nodes run by recognized but not mutually trusting authorities and thus to distribute trust possibly without the complications of a large, diversely trusted network. A single observer, as would threaten a single proxy, will not be effective. Two such well-placed observers or corrupt insiders, however, would be. If one could enforce uniform sending and receiving behavior of a persistent set of users, a stronger adversary might

[6] We ignore herein any attacks other than by compromising or observing network elements, and by altering and observing the timing and/or volumes of traffic sent over network elements and connecting links. For example, any attacks by injecting identifying information into an anonymity circuit or anonymous message are beyond our scope.

be needed. But there is to date no accepted scheme for practically doing this against even a passive adversary. And, if there were, that would not defeat an adversary that selectively blocks or manipulates traffic at sources or destinations. Versions of onion routing that do not use entry guards [34] are statistically broken against a small number of well-placed attackers for longterm repeated connections between the same sources and destinations.

As has already been noted, an adversary observing an entry node and an exit node of the existing Tor network can trivially link the two ends of the connection and correlate source to destination regardless of the number of users of these or any other nodes. Anonymity is broken if the endpoints are owned and not if they are not. There are also potential attacks such as website fingerprinting [22,28] or network latency [23] that require only the entry node to be observed by the adversary. But, in contrast to the entropist view, this again is not significantly affected by the number of other simultaneous circuits initiating at that node and not affected at all by the numbers of circuits elsewhere in the network.

The problem Mixmaster faces against The Man is not the strength of its component protection mechanisms per se but its usability and the implications thereof. (Similarly for Mixminion [9]). With enough compromised nodes it becomes more likely that messages will traverse an entirely compromised path, but many paths might be at most partially compromised. And in mix systems, unlike onion-routing systems, it is not trivial to recognize the same traffic when it is observed in two places. Thus, the number of messages entering and leaving uncompromised mixes does affect uncertainty, and that is indeed the focus of much analysis of mix systems. But there are important factors that largely obviate the protection this might afford.

User configurability and interface play a role [41,13,15] in the amount of use an anonymity system gets, but just as important is the latency inherent in the design: the less the latency the more the system becomes vulnerable to correlation attacks. But, keeping the latency high as in remailer mix networks means that only the users with the most sensitive requirements will use the system [1]. And the most common interactive bidirectional applications such as web browsing, chat, or remote login will not work at all, which is probably an even larger factor limiting adoption. Tor has an estimated quarter million or more concurrent users routing over thousands of network nodes. Though Mixmaster has been around longer, it has had perhaps a few hundred users on any given day, and the network has never been more than a few dozen nodes. Mixmaster's numbers have not grown significantly for years, and the incentive issues we have been discussing imply this to be an inherent limitation. If your goal is to have legal deniability for activity this may be adequate, at least initially before The Man has decided to focus on you. But if your adversary *is* The Man, then avoiding suspicion at all is likely an important goal. As such a few hundred initially possible sources or destinations hidden by a few dozen nodes is just inadequate. Simple externalities such as geographic locations associated with IP addresses, previously known associations, etc. are likely to narrow the initial set even further very quickly [12]. And The Man can own many of the nodes in the

network if he so desires so that the already small number of initially unexposed source-destination pairs shrinks commensurately.

The important thing is that the number of possible targets and number of initially needed network observation/compromise points is small enough to be within the capabilities of The Man to bring other resources to bear irrespective of the ability of the anonymization network to render its users indistinguishable. The issue is not simply a question of uncertainty in identifying a large set of senders and/or recipients, it is an issue of available resources and their expected effectiveness. Yes, The Man strives to reduce his uncertainty about linking senders and recipients, but the measure of his ability to succeed is not the size of those sets or probability distributions on them. Rather it is the resources it will take him to place himself in a position to learn those linking relations. And uncertainty about which elements to attack to improve his position may also play a role, but its role is not as important as the expected cost and benefit of attacking each of them, especially if he has the capability to effectively attack all of them—as is the case for the deployed Mixmaster network.

It is true that entropist approaches might help narrow the initial set of a few hundred senders of a given message. For example, if communication pairs are maintained over hundreds of communication rounds, then statistical disclosure [10,29] may be useful. And The Man should be able to further reduce anonymity within network traffic by DoS attacks on uncompromised nodes [16,4]. If that would be too easily detected, The Man can combine DoS with bridging those nodes via trickling and flooding from his own clients [39] and owned nodes or links. Note that while these active attacks can be analyzed in the entropist model for effectiveness, no entropist view is necessary to simply deploy them in order to slightly improve the efficiency of an already effective compromise. They are just not the attack threats that matter most when going against The Man.

In sum, all current low-latency anonymity systems are broken against The Man. Onion-routing systems by nature of their potential for wide distribution are relatively resistant to other kinds of adversaries. Mix networks, such as Mixmaster and Mixminion, do have measures against The Man. But they are overkill against other expected attackers. This, coupled with the overhead of trying to use them, limits the numbers of both users and infrastructure providers to a point that they are vulnerable to direct inspection and attacks by The Man for which the anonymizing network becomes irrelevant. While entropist techniques can play some role here, they are not the central issue.

5 So What Is Anonymous Communication?

What the entropist conception gets right is that, like other security properties, anonymity protection is determined by the amount of work an adversary needs to do to overcome it. What it gets wrong is the kinds of work the adversary needs to do. We have observed so far that an anonymity breaking adversary generally does not do best by focusing directly on reducing an initially known set of senders (receivers, etc.) to the point that the sender is uniquely determined or has low enough entropy to be considered busted. The entropist approach will

yield theoretically interesting analyses. It will yield system designs that might be useful in controlled settings like elections, where we can control or predict the number and nonidentity of participants and where anonymity within expected-size sets is useful. But the entropist approach is not appropriate for general communication on large diversely shared networks like the internet. I do not set out here a definitive answer of what to replace it with: this is a breaking paper not a design paper. I do, however, offer some hints and suggestions.

As we have noted, putatively more secure designs, such as DC-nets or mix networks, are actually not secure against the intended adversary unless there is a significant increase in both network size and persistent users. And usability and incentive issues make that unlikely. A network might scale up to a point that, even if The Man were able to compromise much communication passing through the network, any individual communication would not be likely to be de-anonymized. P2P anonymity designs have this potential, although they require further scrutiny because they are complex, and attacks on, e.g., network discovery or versions of sybil attacks seem hard to eradicate. Decentralized designs with semicentralized management of network information, such as Tor's, also have this potential. But in either case one is forced to make some assumptions about the likelihood that a given percentage of nodes is under hostile control, and this can be impossible to gauge. Any design that assumes a roughly uniform distribution of resistance to attack across all nodes is likely doomed if The Man can own botnets within the network (and why shouldn't he?). A useful metric should therefore evaluate rather than assume how hard it is for the adversary to own a portion of the network. This much is true of mix networks as well as onion-routing networks and true of peer-to-peer as well as more centralized versions of either of these.

How can we gauge an adversary's difficulty in owning chunks of the network and the traffic flowing over them? We need not analyze how hard it is to contribute hostile servers, break into nodes, steal administrative passwords, control or observe inputs and outputs to nodes etc. For our metrics we just need a representation of the possible outcomes of that analysis. Our observations suggest that it will be useful if those outcomes are not considered uniform across all nodes. So it is not just a question of implementing the strongest component design and deploying it widely. We will also need to consider how much the adversary learns from the nonuniformity itself. We should also consider initial awareness of the network, whether the adversary knows the same network elements and structure as do the communicants he is attacking. Such considerations are a focus of our current ongoing research.

6 Rotor and Other Immodest Proposals

In this paper we have argued that entropist metrics for anonymity are not appropriate to widely shared and deployed anonymity networks. We have observed that they rely on variables for which it is unrealistic to attach values and that would not significantly affect how well anonymity is protected if they could be determined. For less widely deployed networks, whether centralized or distributed,

they provide the most relevant adversary with a set of communicating pairs that he has the resources to deanonymize.

We end with some implications for anonymity research and system design. In the mid nineties (before decent SSL encryption was available in every browser) encrypted traffic was relatively rare on the internet; the mere fact that a message was encrypted made it interesting. Thus, we used to say that if you encrypt your traffic at all, you should make sure you encrypt it well because it is likely to be scrutinized. The same point applies to anonymizing your communication. Unlike choosing good algorithms and increasing keylength for encryption, however, usability and incentive limitations make it unlikely that users and nodes for mix networks will ever increase adequately to exceed the resources of The Man. And even if they did, there is no way to have any meaningful assurance about the distribution of noncollaborating users. More importantly, there is no way to have meaningful assurance about the distribution of noncollaborating network nodes. Our first modest suggestion is thus that existing mix networks for general internet use should simply be abandoned for other than research purposes. They should continue to be studied for there inherent interest. And, they should be used for applications where it is possible to manage and measure the sets of distinct users and anonymity providers and the probability distributions on their behaviors, voting being the clearest example. But for general internet use, they are overkill against almost every adversary except unrealistic ones like the GPA or incredibly strong ones like The Man. And, because of usability and incentive limitations, in practice they do not scale enough to protect against The Man anyway. On the other hand a widely distributed network like Tor may already offer better, though still inadequate, protection. If we modify the design of onion-routing networks like Tor so that trust is incorporated into routing (as we discuss briefly below), then it should be harder for The Man to attack where he must to be effective.

Also, we should not worry (too much) about partitioning attacks. In distributing network directories, Tor has conservatively opted for distributing awareness of the entire directory to clients. This has presented scaling challenges, but the ability to differentiate users by partitioning what they know and don't know is not fully understood. And there have been published epistemic attacks on both mix and onion-routing network designs [8,11]. But unless the adversary is approaching The Man in both capability and intent, this is not a significant concern. And if the adversary is on the order of The Man, then this is not effective. A caveat is that the design should not allow simple actions on the part of a small player, one that can affect at most a small part of the infrastructure, to partition to a point that it can effectively isolate a small number of clients with little effort. Entropy is not useless for characterizing anonymity, it is just not the primary measure of a practical anonymity network's security and as such should not be the primary driver of design.

If the primary measure of anonymity is how hard it is for an adversary to observe or own parts of the network that allow him to conduct attacks, then we can represent the amount of trust we have that a node will not be compromised

and then route accordingly. In addition to anonymity system nodes, the same basic trust notion can be applied to links and destinations. We are currently researching such modeling and analysis of trust and have several initial results for optimal route selections assuming simply that network nodes are trusted at different levels [25]. Our results are based on a partial adversary that can attempt to compromise a fixed fraction of network nodes and succeeds for any given node n_i with probability p_{c_i}. This is a variant of the roving adversary introduced by Ostrovsky and Yung [33] and applied to anonymous routing in [42]. What we add is the idea of trust in a node n_i, represented as $1 - p_{c_i}$ and applied as just described.

We expect the adversary model that we ultimately develop fully will add mobility to the above, much as Ostrovsky and Yung's original work did. This will allow for an iterative attack in which an adversary can discover parts of the network, trust relationships, communication patterns, etc. and then apply what is learned at each stage to conduct further attacks. Anonymity then becomes a question of what an adversary can uncover given a budget of resources expended over an attack of a given duration.

Once better understood, one could imagine a trust-based onion routing network: roTor. This would be a variant of Tor built to counter such an adversary that gets its protection through trust; hence the name.[7]

A rotor network may have a place for mixing too. Routing based on trust could mean that the expected chance that a given communication is compromised is quite low. But, for very sensitive communications it still may pay to have some defense in depth so that observing in a few opportune spots will not yield easy timing correlations. The goal is likely to be hiding of particular linkings rather than avoiding suspicion that the communicants talk at all. But that is likely to yield rather different designs from existing mixes and mix networks, for example, embedding messages to be mixed by an intermediate node inside of web traffic on a low-latency circuit.

This last section has been fairly speculative so as to attempt to give at least one vision of what definitions and system designs might replace entropism. Whether or not these ultimately prevail, however, is not the central point of

[7] Tor began as *the* onion routing, designed as one of the original NRL onion routing projects, in contrast to the many other onion routing designs that followed the NRL original. It is also a recursive acronym for *Tor's onion routing*. ('Tor' has never been an acronym for *The Onion Router*—as generally misstated in the press, nor has it ever been proper to write it 'TOR'.) Though that is perhaps enough of a hint to make the following implicitly obvious to the reader, we note that 'rotor' is a sesquipalendromic sesquirecursive acronym for *rotor onion T*^{3}*or's onion rotor*, which itself expands into *rotor onion [Tor's trusted through] onion rotor*. As also implied by the name, the T-words rotate so that it also expands to *rotor onion [trusted through Tor's] onion rotor* and to *rotor onion [through Tor's trusted] onion rotor*, all three of which should be viewed as in the same equivalence class. Defining the algebraic structure for the equivalence relation is left as an exercise for the reader. Note also that the nodes of a rotor network are the network rotors; thus a rotor network *in sensu composito* is also a rotor network *in sensu diviso*.

this paper. The central point is that the model on which all existing work on anonymity, the entropist model, is broken for open or widely shared communications. That result remains, even if we do not know definitively what to put in its place.

None of the usability, incentive, or usage observations made herein are particularly novel. However, engendered by its fundamental entropism, the community still operates on a premise that mix networks are more secure. Those of us who have made such observations have even sometimes moderated our assumptions so that mix networks are described as more secure in principle if not in current deployment. But taking the next steps implied by these observations has not even been overtly broached. All of us have still clung to an entropist view of anonymity and thus to the systems and analysis it engenders. Hopefully, this paper will give us the push we need to let go.

Wait, what about interference attacks that allow remote observation without compromising or directly observing nodes and links [32,31,18]? Sorry. We've got an answer, but there's no room in the margins. See the next paper.

Acknowledgments. I thank Claudia Diaz, George Danezis, Roger Dingledine, Aaron Johnson, and Nick Mathewson for helpful discussions that led to better articulations of the ideas in this paper. (Nonetheless any remaining incoherence is my own.) This work supported by ONR.

References

1. Acquisti, A., Dingledine, R., Syverson, P.: On the Economics of Anonymity. In: Wright, R.N. (ed.) FC 2003. LNCS, vol. 2742, pp. 84–102. Springer, Heidelberg (2003)
2. The Anonymizer (2009), `http://www.anonymizer.com/`; Homepage of the company that offers the Anonymizer Proxy Service. Original Anonymizer first described in [5]
3. Berthold, O., Federrath, H., Köpsell, S.: Web MIXes: A System for Anonymous and Unobservable Internet Access. In: Federrath, H. (ed.) Anonymity 2000. LNCS, vol. 2009, pp. 115–129. Springer, Heidelberg (2001)
4. Borisov, N., Danezis, G., Mittal, P., Tabriz, P.: Denial of service or denial of security? How attacks on reliability can compromise anonymity. In: De Capitani di Vimercati, S., Syverson, P., Evans, D. (eds.) CCS 2007: Proceedings of the 14th ACM Conference on Computer and Communications Security, pp. 92–102. ACM Press (2007)
5. Chaum, D.: Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM 4(2), 84–88 (1981)
6. Chaum, D.: Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM 4(2), 84–88 (1981)
7. Clauß, S., Schiffner, S.: Structuring anonymity networks. In: Goto, A. (ed.) DIM 2006: Proceedings of the 2006 ACM Workshop on Digital Identity Management, Alexandria, VA, USA, pp. 55–62. ACM Press (2006)
8. Danezis, G., Clayton, R.: Route fingerprinting in anonymous communications. In: Sixth IEEE International Conference on Peer-to-Peer Computing, P2P 2006, pp. 69–72. IEEE Computer Society Press (2006)

9. Danezis, G., Dingledine, R., Mathewson, N.: Mixminion: Design of a type III anonymous remailer protocol. In: Proceedings of the 2003 IEEE Symposium on Security and Privacy, Berkeley, CA, pp. 2–15. IEEE Computer Society (May 2003)
10. Danezis, G., Serjantov, A.: Statistical Disclosure or Intersection Attacks on Anonymity Systems. In: Fridrich, J. (ed.) IH 2004. LNCS, vol. 3200, pp. 293–308. Springer, Heidelberg (2004)
11. Danezis, G., Syverson, P.: Bridging and Fingerprinting: Epistemic Attacks on Route Selection. In: Borisov, N., Goldberg, I. (eds.) PETS 2008. LNCS, vol. 5134, pp. 151–166. Springer, Heidelberg (2008)
12. Danezis, G., Wittneben, B.: The economics of mass surveillance and the questionable value of anonymous communications. In: Anderson, R. (ed.) Fifth Workshop on the Economics of Information Security, WEIS 2006 (June 2006)
13. Dingledine, R., Mathewson, N.: Anonymity loves company: Usability and the network effect. In: Anderson, R. (ed.) Fifth Workshop on the Economics of Information Security, WEIS 2006 (June 2006)
14. Dingledine, R., Mathewson, N., Syverson, P.: Tor: The second-generation onion router. In: Proceedings of the 13th USENIX Security Symposium, pp. 303–319. USENIX Association (August 2004)
15. Dingledine, R., Mathewson, N., Syverson, P.: Deploying low-latency anonymity: Design challenges and social factors. IEEE Security & Privacy 5(5), 83–87 (2007)
16. Dingledine, R., Syverson, P.: Synchronous Batching: From Cascades to Free Routes. In: Martin, D., Serjantov, A. (eds.) PET 2004. LNCS, vol. 3424, pp. 186–206. Springer, Heidelberg (2005)
17. Dolev, D., Yao, A.C.: On the security of public-key protocols. IEEE Transactions on Information Theory 2(29), 198–208 (1983)
18. Evans, N.S., Dingledine, R., Grothoff, C.: A practical congestion attack on Tor using long paths. In: Proceedings of the 18th USENIX Security Symposium, Montreal, Canada, pp. 33–50. USENIX Association (August 2009)
19. Goldschlag, D.M., Reed, M.G., Syverson, P.F.: Hiding Routing Information. In: Anderson, R. (ed.) IH 1996. LNCS, vol. 1174, pp. 137–150. Springer, Heidelberg (1996)
20. Goldschlag, D.M., Stubblebine, S.G., Syverson, P.F.: Temporarily hidden bit commitment and lottery applications. International Journal of Information Security 9(1), 33–50 (2010)
21. Helmers, S.: A brief history of anon.penet.fi - the legendary anonymous remailer. CMC Magazine (September 1997)
22. Hintz, A.: Fingerprinting Websites Using Traffic Analysis. In: Dingledine, R., Syverson, P. (eds.) PET 2002. LNCS, vol. 2482, pp. 171–178. Springer, Heidelberg (2003)
23. Hopper, N., Vasserman, E.Y., Chan-Tin, E.: How much anonymity does network latency leak? In: De Capitani di Vimercati, S., Syverson, P., Evans, D. (eds.) CCS 2007: Proceedings of the 14th ACM Conference on Computer and Communications Security, pp. 82–91. ACM Press (2007); Expanded and revised version in [24]
24. Johnson, A., Syverson, P., Dingledine, R., Mathewson, N.: Trustbased anonymous communication: Adversary models and routing algorithms. In: CCS 2011: Proceedings of the 18th ACM Conference on Computer and Communications Security, ACM Press (October 2011)
25. Johnson, A., Syverson, P.: More anonymous onion routing through trust. In: 22nd IEEE Computer Security Foundations Symposium, CSF 2009, Port Jefferson, New York, USA, pp. 3–12. IEEE Computer Society (July 2009)

26. Johnson, A., Syverson, P., Dingledine, R., Mathewson, N.: Trust-based anonymous communication: Adversary models and routing algorithms. In: CCS 2011: Proceedings of the 18th ACM Conference on Computer and Communications Security. ACM Press (October 2011)
27. JonDonym – the internet anonymisation service (2008), `https://www.jondos.de/en/`; Commercial version of the qJava Anon Proxy (JAP). Initially published description in [3]
28. Liberatore, M., Levine, B.N.: Inferring the source of encrypted HTTP connections. In: Wright, R.N., De Capitani di Vimercati, S., Shmatikov, V. (eds.) CCS 2006: Proceedings of the 13th ACM Conference on Computer and Communications Security, pp. 255–263. ACM Press (2006)
29. Mathewson, N., Dingledine, R.: Practical Traffic Analysis: Extending and Resisting Statistical Disclosure. In: Martin, D., Serjantov, A. (eds.) PET 2004. LNCS, vol. 3424, pp. 17–34. Springer, Heidelberg (2005)
30. Möller, U., Cottrell, L., Palfrader, P., Sassaman, L.: Mixmaster protocol - version 3. IETF Internet Draft (2003)
31. Murdoch, S.J.: Hot or not: Revealing hidden services by their clock skew. In: Wright, R.N., De Capitani di Vimercati, S., Shmatikov, V. (eds.) CCS 2006: Proceedings of the 13th ACM Conference on Computer and Communications Security, pp. 27–36. ACM Press (2006)
32. Murdoch, S.J., Danezis, G.: Low-cost traffic analysis of Tor. In: Proceedings of the 2005 IEEE Symposium on Security and Privacy, IEEE S&P 2005, pp. 183–195. IEEE CS (May 2005)
33. Ostrovsky, R., Yung, M.: How to withstand mobile virus attacks. In: Proceedings of the Tenth ACM Symposium on Principles of Distributed Computing, PODC 1991, pp. 51–59. ACM Press (1991)
34. Øverlier, L., Syverson, P.: Locating hidden servers. In: Proceedings of the 2006 IEEE Symposium on Security and Privacy, S&P 2006, pp. 100–114. IEEE CS (May 2006)
35. Parekh, S.: Prospects for remailers: where is anonymity heading on the internet? First Monday 1(2) (August 5, 1996), `http://www.firstmonday.dk/issues/issue2/remailers/`
36. Serjantov, A., Dingledine, R., Syverson, P.: From a Trickle to a Flood: Active Attacks on Several Mix Types. In: Petitcolas, F.A.P. (ed.) IH 2002. LNCS, vol. 2578, pp. 36–52. Springer, Heidelberg (2003)
37. Pfitzmann, A., Köhntopp, M.: A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, version v0.32 (December 2009), `http://dud.inf.tu-dresden.de/Anon_Terminology.shtml`, Regularly revised and updated version of [36]
38. Reed, M.G., Syverson, P.F., Goldschlag, D.M.: Anonymous connections and onion routing. IEEE Journal on Selected Areas in Communications 16(4), 482–494 (1998)
39. Serjantov, A., Dingledine, R., Syverson, P.: From a Trickle to a Flood: Active Attacks on Several Mix Types. In: Petitcolas, F.A.P. (ed.) IH 2002. LNCS, vol. 2578, pp. 36–52. Springer, Heidelberg (2003)
40. Simmons, G.J.: The history of subliminal channels. IEEE Journal on Selected Areas in Communications 16(4), 452–462 (1998)
41. Syverson, P., Reed, M., Goldschlag, D.: Onion Routing access configurations. In: Proceedings DARPA Information Survivability Conference & Exposition, DISCEX 2000, vol. 1, pp. 34–40. IEEE CS Press (1999)

42. Syverson, P., Tsudik, G., Reed, M., Landwehr, C.: Towards an Analysis of Onion Routing Security. In: Federrath, H. (ed.) Anonymity 2000. LNCS, vol. 2009, pp. 96–114. Springer, Heidelberg (2001)
43. Tang, C., Goldberg, I.: An improved algorithm for Tor circuit scheduling. Technical Report CACR 2010-06, University of Waterloo, Center for Applied Cryptography Research (2010), http://www.cacr.math.uwaterloo.ca/techreports/2010/cacr2010-06.pdf
44. The Tor Project (2009), https://www.torproject.org/; Homepage of the non-profit organization that maintains and develops the Tor network. Original Tor design first published in [14]

Why I'm Not an Entropist
(Transcript of Discussion)

Paul Syverson

Naval Research Laboratory

The previous talk[1] was about trying to get entropy, and I'm going to talk about ignoring entropy. It's natural to talk about anonymity in a workshop about brief encounters and security protocols, I would say anonymity is the quintessence of brief encounter, you only have a brief encounter if in fact you are anonymous, so you need to be anonymous to guarantee that it is brief, otherwise it is at best pseudonymous, because you're preserving state from one instance of communication to another.

Now this is the standard definition of anonymity that's been around for a long time, anonymity is when you've got a bunch of Alices in some set and the attacker can't figure out, can't distinguish, which of them is talking to Bob. Now can't distinguish might mean that you have a set of some size and that's the measure of indistinguishability, or you might have some kind of probability distribution. This is what I call the entropist conception. I don't want to get too hung up on the term entropy, I thought it was reasonable because it ties up with a lot of different notions. There's lots of different notions of entropy: basic anonymity set sizes, Hartley entropy, everybody uses Shannon entropy, some people feel that min entropy is more appropriate for some things.

I don't really care too much about the specific mathematical characterisations, and I think there are lots of places where entropy is quite useful. What I want to question, is the notion of entropy where you just have some set — and typically we pick on the senders as the quintessential example, but it could be any number of things, Pfitzmann and Hanson[2] called it items of interest — and that you have some kind of uncertainty in that set, on the part of the adversary. The reason this is important is because this is the notion of anonymity that underlies virtually all system design, all theory, all research. Everything in anonymity basically has this as its starting point, and the reason I want to talk about it is because I think we've just been doing it wrong for thirty years.

Now, I should qualify what I mean by that. I think there are plenty of places where this is a very useful and right notion. We were talking about voting earlier this morning, I think that's a perfectly reasonable place to look at entropy as a measure of anonymity. But my focus is in looking at large-scale internet type communications: web browsing, or sending email to people who may not know anything about your anonymity system but you still have to communicate with

[1] Bonneau, these proceedings.

[2] "Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology" by Andreas Pfitzmann and Marit Köhntopp in LNCS 2009: Designing Privacy Enhancing Technologies.

B. Christianson et al. (Eds.): Security Protocols 2009, LNCS 7028, pp. 231–239, 2013.

them, these are standard uses for which the paranoid cyberpunk sort of people think, oh, I've got to have this kind of thing, and I think that's wrong, this is not how to do it.

For one, it may not be meaningful to even stipulate some unique set, there may not be a set on which you can attach any measure in any reasonable way. I would say that what you want to ask is, how hard is it for some attacker to break the anonymity of the system, and that's going to be your anonymity metric. I'll come back later to talking about the steps, where the details are, how do you decide what hard means. But the important thing is that I will argue (this is mostly a breaking not a constructing paper), is that the entropy approach, the entropist's approach, does not provide a good measure of anonymity for these sorts of systems. So just in case I was being too gentle, my goal is to fundamentally remake our understanding of anonymity for widely used Internet communication systems.

These are not all the anonymity systems out there for communication, there's tons of them, but these are some that have had the most wide use in the last many years. They're either centralised or distributed, I'm going to ignore the centralised ones, they have single points of failure and other issues, and I'm going to focus on the distributed ones. You basically have two broad types, most of these are some flavour of onion routing, and then the other alternative, for high-latency systems, is essentially some kind of mix network. Onion routers are not mixes at all, people always think they're based on it, it's a popular misconception, but I can tell you since I came up with it, that onion routing was not at all based on mixes. Anyway, the reason that we make this distinction between low and high-latency is, the low-latency systems are totally vulnerable to correlation attacks: because it has to be low-latency you can just look at the pattern of packets and messages as they go through the system, and you can just watch the pattern and you can break it. If you've got more latency, as you do in a mix system, then you can start mucking around with the order and the timing, and you can make it much harder to do this. This has been confirmed over and over again in experimentation, and in analysis that basically there's nothing that beats this against any kind of reasonable attacker, usually anything that's proposed is hugely expensive, and it doesn't work anyway, so this is just something you have to live with.

So the prevailing wisdom is that the high-latency systems are more secure, because you don't have this correlation attack, and it also complicates some other kind of attacks as well. The trade-off is they can't be used for interactive things, low-latency things, you can't do web browsing over them, you can't do remote login, you can't do a lot of things that people like to do. So my question is, if this is supposed to be the more secure system, what is it more secure against?

In order to think about that you have to say, well what is your adversary, so the question then is, what is a reasonable adversary when you're talking about a practical anonymous internet communication system. There's lots of different possible things you might be worried about, but I think they're pretty much subsumed by one adversary that you can focus on, and that's The Man.

The Man owns huge chunks of the anonymity infrastructure because he's either compromised them, or he's providing it, you know Tor is volunteer networking, he can just volunteer; so is Mixmaster. He certainly is going to have access to ISPs, backbones, what have you servers. About individuals he's going to know answers, maybe not to all the questions we were just hearing about, but to the things that you can find on Facebook, he's going to know that, and maybe some other stuff as well, and if he cares about you he is going to tap your phone, he's going to follow you around.

So you should be thinking of The Man as on the level of, state level actors, intelligence organisations, secret police; it could be organised crime, maybe a large corporation if they're going up against their adversary or something. This is the adversary that your basic cyberpunk is convinced is after them. So the question is, how well do you actually do against The Man.

Now it's important to keep in mind he's really big, he's really powerful, but he's not global, and he's not omnipotent.

Matt Blaze: But that's just what you want us to think, because you're part of The Man.

Reply: I'm not at liberty to discuss that on the record, but if you want to take that offline I know a pub where we can, yes.

Ben Laurie: The Man has just asked me to point out that Google is global and omnipotent.

Reply: I would contend not, but thank you for pointing out that I was right to bring corporations into that. If you're not worried about The Man then certainly doing mixing is just overkill, right, because if you're not worried about somebody who's watching all these things everywhere then you don't need this level of protection. But the flipside is that if you are worried about The Man, if you're worried about him suspecting you and keeping an eye on you because he thinks you're doing bad things, mix networks aren't going to help you either.

That's because they don't scale, and I'll come back to that point, but I would like to dismiss one property that I think mix networks do provide reasonably well, and that's plausible deniability[3]. Even the mix nets that are in wide use on the internet today, have maybe a couple of hundred users at most on any given day using them, that's probably generous, and there's maybe a dozen or two dozen nodes in the network at peak, but I would contend that the deniability you get in this setting is basically irrelevant. This is because you're playing a shell game, and The Man just isn't going to care because, 50 or 100 people is basically nothing, he doesn't care where the pea is, he's just going to watch all of you. Maybe he doesn't know that you sent this particular message, but if your goal is to hide the fact that you associate with somebody, once he's got a set of maybe 50 or 100 people, right, that's enough, that's just a tiny set, you just watch all 50 or 100 and game over.

[3] See "Cryptography and evidence", Michael Roe, University of Cambridge PhD dissertation, available as `http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-780.pdf`.

Now you might get out of this if you could just build a big mix network, because as I said, he's not global and he's not omnipotent, and they could scale in principle, but they can't, they haven't, and they won't. This has to do with usability and incentives, because most people are quite correctly not at all worried about The Man. There's very few people who would be right to worry about The Man, and most of the people who do worry about The Man are not those people. People are worried about identity thieves, or abusive ex spouses or whatnot, and for these purposes a distributed low-latency system is going to be more than adequate. Mixmaster, you've got maybe 200 users a day protected by a few dozen mixes, so your entropy is maybe 7, maybe 8 bits, and that's your original entropy, assuming that The Man isn't applying any exogenous information, which he probably is. If you're something like Al Qaeda, he's probably got a pretty good idea who's in Al Qaeda and who isn't, even though he doesn't have it exactly. But if you're going to be hiding, of course he can own several of the mixes, plus he's going to be able to see all the communication certainly on all the links, but even within some of these mixes they're not going to be doing anything. So the bottom line is, with like 200 people maybe initially, you can expect that he's not going to really have to watch those 200 people, he's going to have to watch maybe 20.

Now for Tor circuits it gets a little more complicated. I'm assuming I don't have to explain what a mix is, or what Tor is.

Bruce Christianson: You're amongst friends here Paul.

Reply: It's a mix, you get a bunch of stuff, you throw it in, you shake it up, and then when it comes out you can't tell which thing that came out, was what thing came in there, that's a mix. So for Tor, I would say that this point is even more clear, that it's hard to say what the set is because who's using it from one day to the next is going to vary, so the set you saw yesterday may have nothing to do with the set you saw today, whereas if you watch the mix network for a day you're going to know everybody who sent into it and out of it, for a day you'll know everybody who sent that day. George is saying, no, how could you not know?

George Danezis: Because a lot of messages will have actually originated from the previous day, and will be in queues, and a lot of messages that came in on that day will go out the next. It is actually a distribution along a huge set.

Reply: But the expected time that a message sits in a mix now is not all that long, and again, it's this trade-off, I mean, the longer it sits the smaller the user set.

George Danezis: But anonymity sets are defined on basically infinite populations because there is a possibility your message comes up in a million days. This is at the core of your argument, I think.

Reply: This is back to the plausible deniability thing, you watch it for a month, the probability that a message that was sent in was from before that month is very small, you're still only looking at 200 people.

Bruce Christianson: Paul's argument is that the odds favour The Man.

Joseph Bonneau: Well if the question is what holds up in court, I think that evidence ...

Reply: No actually, I'm ignoring that, that was my point about the shell game. This is more like your favourite Schwarznegger movie, you know, where you have the über bad guy who has some vague Eastern European accent, right, and you have the two like naïve stick figures from xkcd standing there passing something behind their back, and they're like, ha ha, you don't know which of us has the object you want, and he just shoots them both. The Man doesn't care. Now there are people that some people would call The Man, and he does care, he's going to follow due process and stuff, but I'm just contending that the size of the number of people amongst whom you're hiding is so small that The Man is just going to watch all of you. If you want to hide the fact that I sent the message with this content, yes, it will work, and that's OK, but most of the people are not using it for that purpose, they don't want The Man to know that I might be the guy who's sending to this other guy, and there's just so few people that you're looking at, it doesn't work.

Now, with Tor, I think you're actually in some sense better off against The Man in that, because it's so big and distributed, you have some statistical chance of not getting watched. But I did a paper with Lasse Overlier back in 2006[4] in which we found that if you could find only a single Tor node, you could find a hidden server in the Tor network in minutes. This was a surprise. We knew in principle you could do it, we didn't know you could do it that quick and easy, that network is a little smaller, etc, etc, but this was one of the things that made us change the way we decided to handle connecting up to the network. Instead of saying, well I'll just connect up a completely random route, the point was, statistically you get owned fast. Many people don't realise, because onion routers are not mixers and because of the end-to-end correlation attack, if you owned both ends of the communication, then you know who's talking to whom, it doesn't matter what happened in the middle. Basically we don't try to solve that problem, we just live with it, but if statistically, if you're going to open up a bunch of connections you're going to be found pretty fast, so just accept that and pick one or a small number of starting points, hope you picked right the first time, and that they're not compromised, if they're compromised all of your traffic is compromised, but you were pretty badly off anyway, this way if none of them are compromised then you will never be owned, so that's nice. But of course never isn't never, because sooner or later things go down and stuff, and you have to rotate, then get new Guard Nodes and so on, and because of that, eventually you get caught by The Man, so Tor's not going to help you either.

So what can you do? The main point of my talk was that you're screwed, but I'm not going to end on that pessimistic note, I'll have another one later. What I think the entropist gets right is that anonymity, just like any other security

[4] "Locating Hidden Servers", in: Proceedings of the 2006 IEEE Symposium on Security and Privacy.

property, is about how hard is it for the adversary to break what you have, and what the entropist conception gets wrong is that the measure of the work is not how hard it is to narrow down a set. Even though if you are breaking anonymity you can find a set and you can show that, when you're done breaking it, there's a smaller set than you started with. One issue is that it's not necessarily a subset, that's one way in which entropism breaks, but more importantly, I just don't think that's how what's going on is happening, that's just like an epiphenomenon that you notice afterwards.

So here's some suggestions, and I don't claim that I've got it right now. One thing you could do for the metric is just consider how hard is it to gain access to all the things that The Man needs to own. I'm not really sure how you want to measure this, one thing you might just say is, give The Man a budget in dollars, and for example, to own a node, ask what's the cheapest way? It might be to bribe a janitor, it might be to get a server in the same facility and then if you can somehow have access to that, maybe you actually have to do some crypto breaking things, so how many FPGAs do you have to get, and how much does it cost to program it. And then your metric for anonymity is just going to be the expected value of dollars and time to do the break.

George Danezis: I'm worried that you're going to finish on a happy note.

Reply: Oh no, I've got a couple of more slides.

George Danezis: In order to get a figure in dollars with the expected cost of actually getting someone, OK, that effectively assumes that you have a probability model that takes the adversary model, the adversary capabilities and can actually cost them as well, and also the system, the observation and all that stuff that are the result of the adversary's capabilities, and then effectively do an analysis that tells you, what is the probability of getting the right people, and then find out how much they need to invest until they nab someone.

Now my argument is that this is actually a more general process than the process of taking effectively a probabilistic model assuming a particular adversary model, and then concluding probability distribution then entropy. So effectively what you're saying is we need to do more in order to have a more complete picture of how good the anonymity we get is in practice, and for realistic adversaries. But my argument is that in order to do the costing, the risk management, the risk analysis, and all the costing you'd have to do, you will do exactly the same process mathematically as calculating the entropies. You will have to have a probabilistic model of the observations.

Reply: You know what asset you think you're protecting, so you can say, I know what I'm hiding behind, and this is the usual difference between vulnerability and threat, I know what I've got and how it's protected, and then I can then consider, it's a separate thing whether or not the adversary knows, for example, that this piece is behind that piece. Now, you know, he might just jump past some of the pieces, but I can still effectively think about it iteratively, and I can say, look, once the adversary breaks this, then he's going to learn more, he's going to figure out, oh now I can try to do these other things, but at no point

does he have some sort of single set that he's trying to somehow narrow down, he's just gaining more information. I'm not claiming that I actually, know what we should have, this is the part where I think I'm weakest, but I'm just saying that what we did have is pie in the sky.

Ross Anderson: There's a very interesting distinction here between capital cost and marginal cost for The Man. The Man here in Britain at present pays an ISP £500 to get some traffic data off somebody, and presumably a similar fee to put a black box on a LAN, and if you're a small ISP you have got the black box already. There is a measure to centralise all traffic data in a big database in Cheltenham, and having spent £15 billion on that, The Man's marginal costs for future enquiries is zero. Now given the way that economics works with big capital costs and low marginal costs, there are things we can say about this, they're subtle and they're non-linear.

Reply: Yes, but I would say that The Man has not compromised your anonymity unless you're going to and from locations in Britain entirely in your communication, and he's willing to analyse all that data to look for the correlations of the timing and such.

Ross Anderson: But if The Man has got a black box in every ISP in Britain, and assume that The Man can download applets into the black boxes, then there are various distributed correlation algorithms that it might be quite fun for an engineer to design.

Reply: Right, but already right upfront he's only going to win on the things he can see, so if you don't happen to be seen, then he wins on one end of the communication but he doesn't necessarily win on the other in that case, it's a jurisdictional arbitrage point.

George Danezis: Paul, I think you're overly worried about the fact that the adversary does not see part of the stuff that is going on in a big population of users, let's say like in Tor. But as Donald Rumsfeld said, there are known knowns, known unknowns, and unknown unknowns, and to some extent this is a known unknown, in that I know as an adversary that I can intercept 10% of the communications, and therefore I know that roughly 90% of the users I cannot see. That does not mean that I cannot have a good probabilistic model of what is my probability out of the people I see, is it someone communicating with someone else, or is it just that I don't see it at all. And actually you need to have an estimate of this figure in order to even cost how expensive this is.

Reply: But you're assuming something that I reject, which is that the goal of The Man is to see as much as he can.

George Danezis: The goal of The Man is to detect someone.

Reply: To detect something particular, he's not trying to just hoover things up.

Sandy Clark: Paul, I disagree because of the incident in San Francisco where they slipped the fibre into a box-type trunk, they were just collecting masses of data to see what they could find for a one time expense.

Reply: Yes, right, OK. That's perfectly compatible with what I'm saying, that would go into your evaluation of what it costs. The point is, any communication going through that part is going to be whatever it costs them to set that up, that's it. But I think that largely you are screwed against The Man, the only way forward, to the extent that there's something to go forward on, is going to be once you bring trust into ...

George Danezis: How do you measure how screwed you are against The Man here? I think that's what we're nitpicking about, what is your measure of how bad things are?

Reply: Well tell me who The Man is first of all, because I don't think there is just one, The Man, there's going to be different ones. If you're British Intelligence I think you're probably not as worried about The Man that Ross just told us about, as you are about certain Other Men, right. I think that what you need to do, a priority that you might have, is if you can bring some sort of notion of trust into this. I have a forthcoming paperwhere we look at, suppose you do actually have paths.

This is a preliminary result, just looking at the nodes in the network itself. Right now there's no reason not to think that some large intelligence organisations own 10%, 20% of the nodes in the Tor network. I don't know that they do, but I don't know that they don't, so if you're using the Tor network and if your adversary is for the moment limited to the nodes themselves, that's still a big problem. But if you have some sort of trust metric you can put on the nodes, you don't want to reduce yourself to the initial problem we were trying to solve when we set up onion routing, which was, we can't make this a Navy only system, because anything that pops out of it will be known to come from the Navy, you still have to share the thing, but you still want to have a trusted subset, so how do you do smart routing so you're not revealing that but you're still picking things that are more trusted in an effective way. And we have some optimal route selection results that are forthcoming on that, that's the theory side.

I assume a lot of you know that Tor stands for Tor's onion routing, that's where the name comes from. A lot of people write the onion router, that's completely wrong, that's something somebody in the press messed up, but oh well. Anyway, as I tried to argue, I think for The Man it's already more secure than Mixmaster because at any given point The Man is going to see all the communication going into and out of Mixmaster, going back for some time, and he won't for Tor, so there's at least some sense in which you're better off. But you're still not safe against The Man, and I think you need some sort of notion of trust built in. I think that once you add trust to something like Tor, you have something that I call RoTor, this stands for Rotor onion Tor is trusted to onion router, which I think is the world's first sesqui-palindromic sesqui-recursive acronym,

because of course Tor is also an acronym, and it's palandromic both in the acronym and in the wording, but of course, if this is an equivalence class that you can rotate, it's a rotor right, so you rotate through, so it's all the same.

Matt Blaze: It is currently recursive?

Reply: Yes. If you add trust in then I think you have at least a chance to beat The Man. So in conclusion, the entropist approach doesn't tell you anything really useful and interesting about how good you are against the adversary that prompted the designs in the first place. I think we need to fundamentally rethink the theory underlying this. And in particular the mix networks, which were really designed against The Man; you should stop fooling yourself.

Deriving Ephemeral Authentication Using Channel Axioms

Dusko Pavlovic[1,⋆] and Catherine Meadows[2]

[1] University of Oxford, Department of Computer Science, and
Universiteit Twente, EWI/DIES
dusko@cs.ox.ac.uk
[2] Naval Research Laboratory, Code 5543, Washington, DC 20375
catherine.meadows@nrl.navy.mil

Abstract. As computing and computer networks become more and more intertwined with our daily lives, the need to develop flexible and on-the-fly methods for authenticating people and their devices to each other has become increasingly pressing. Traditional methods for providing authentication have relied on very weak assumptions about communication channels, and very strong assumptions about secrecy and the availability of trusted authorities. The resulting protocols rely on infrastructures such as shared secrets and public key hierarchies that are too rigid to support the type of flexible ad-hoc communication we are growing accustomed to and beginning to rely upon.

Recently, different families of protocols allow us to weaken assumptions about trusted infrastructure by strengthening the assumptions about communication channels. Examples include proximity verification protocols, that rely, for example, on the round trip time of a challenge and response; and bootstrapping protocols that rely upon human-verifiable channels, that is, low-bandwidth communication between humans. The problem now becomes: *How do we ensure that the protocols are achieve their security goals?* A vast amount of literature exists on the formal analysis of cryptographic protocols, and mathematical foundations of protocol correctness, but almost all of it relies upon the standard assumptions about the channels in end-to-end, and so its usefulness for nonstandard channels in pervasive networks is limited. In this paper, we present some initial results of an effort towards a formalizing the reasoning about the security of protocols over nonstandard channels.

1 Introduction

Pervasive computing has become a reality. We have long been used to the idea that computers are everywhere, and that we interact with multiple devices that can interact with each other and with the Internet. But there is another important aspect of pervasive computing. Not only has the concept of a computer and a computer network changed, but the notion of a communication channel is changing as well. Wireless channels, of course, have been a common part of computer networks for some time. Quantum channels are appearing on the horizon. But what is really interesting is the way the nature

⋆ Supported by ONR. Current address: University of London, Royal Holloway, Department of Mathematics/ISG.

B. Christianson et al. (Eds.): Security Protocols 2009, LNCS 7028, pp. 240–261, 2013.