Measuring anonymity: a few thoughts and a differentially private bound

George Danezis

15 May 2013

1 Early Measures

Measuring the quality of protection offered by security mechanisms, including anonymous channels, is of paramount importance. It allows architects to compare designs, and reliably improve them. It also allows them to understand the impact improvements in other aspects of a design, for reliability, performance or usability, may have on the overall security of systems. In the absence of security measures, features that have a demonstrable impact, such as performance, will be prioritised and the security of the system may be inadvertently undermined.

How to measure the anonymity of large deployed system, has been for a long time, and still is an open problem. This is the case despite a surge of interest in the topic shortly after 2002, when entropy based measures were proposed [41, 15]. The most popular early measure of anonymity was the "anonymity set size" [39], a design parameter of a system that roughly denotes amongst how many possible actors an action is hidden, or conversely how many possible actions a specific actor may have performed. The downside of this measure is its lack of subtlety in taking into account the relative likelihood of the relationships. This was first noted by Dogan Kestogan, when analyzing sg-mixes [29], which are implemented using inherently probabilistic mechanisms, and do not produce "clean" anonymity uniformly likely anonymity sets.

Another early measure of anonymity is used in the Crowds system [40], to compute the probability an observed action was performed by a specific actor. Interestingly, Crowds threat model is very weak and might be an inspiration for metrics relating to onion routing as we will see. With hindsight the quantitative element of the measure was very influential, while its qualitative interpretation seems rather arbitrary.

2 Entropy Measures

At PETS 2002 Diaz et al. [15] and Serjantov and Danezis [41] concurrently suggested that information theoretic measures, such as entropy or normalized entropy, may be a good measure for anonymity. This is in line with uses of entropy to compute the security of symmetric key primitives, and its meaning denoting uncertainty about a measured quantity. When applied to the belief of an adversary over who performed an action, or which action was performed it is a summary of their uncertainty.

One of the reasons these works became influential is the use of the metrics to reason about systems that could not be measured accurately before using anonymity sets, namely pool mixes [12] and Crowds [40]. A simple composition theorem was also provided to combine primitive anonymity building blocks – as long as all actors or actions were distinct. This proves to be a very serious limitation. The use of pool mixes as examples to illustrate the metric spurred a long line of research relating to measuring the anonymity of different mixing strategies [12, 14, 43], with and without dummy messages [13], and their susceptibility to blending and $(n-1)$ attacks [42]. To some extent this is a bit of a "red herring", since any strategy with a bounded mean latency will have an upper bound on uncer-

tainty, and probing attacks by the adversary, in what is inherently an open system, defeat most attempts to protect single mixes from blending attacks [36].

The practice of using Entropy based measures suffered over the next 10 years from three fundamental difficulties:

- Shannon Entropy is inherently an average measure. It is not very good at providing intuition about the worse case. Shmatikov first pointed this out, and suggested that min-entropy may be a better measure as a result [44]. This proposal is important, but still suffers from the next two difficulties.

- For the measure to make any sense, one needs to compute the Shannon entropy of the posterior distribution of actions or actors. This includes the prior belief of the adversary and all information they can infer from observing the operation of the anonymity system. It turns out this is very hard for complex systems, and furthermore it only informs the adversary about the quality of one particular run of a system. Getting a lower bound for any possible or likely attacker observation is very hard indeed.

- Finally, it is rather uninteresting to reason about the existence or absence of single interactions between a target actor and actions. Anonymous channels need to provide a robust service to higher level applications that will use them for long term, systematic, persistent patterns of communication. Thus it is imperative for a measure to provide good intuitions about the security of repeated uses of the channel. The entropy measures are very poor at this.

3 Getting to the Posterior

Real anonymous communication channels are complex engineering artefacts. As a result they are both constrained in a variety of ways, and they inevitably leak information about their internal workings, including their handling of secrets, to various classes of adversaries. The task of processing this information to define proper posterior distributions over actions or actors, is necessary to apply any entropy metrics in a sound manner. Interestingly, this distribution is also necessary for measuring the anonymity of the system using the Bayesian probability of error of the adversary [5], or even mutual information metrics [4, 45].

Deriving those for various systems, took a very long time and uses advanced techniques in Bayesian inference. This has led to some notable results, but with important limitations:

- A full Bayesian treatment of Crowds, under its original very weak threat model, and for a single observation of the adversary (again as in the original work), is a success story [7]. One can derive an analytic form for the posterior belief of the adversary; this posterior can be bound to show a minimum over any observation; and finally this minimum can be used to demonstrate that the distribution of latency imposed by Crowds provides the optimal anonymity for a certain mean latency. This is very exciting, if not for the fact that no one considers implementing Crowds, due to its susceptibility to long terms attacks [50] (related to its very weak threat model).

- A Bayesian treatment of a complex network of perfect threshold mixes has provided a simulator that takes traces from an adversary and produces distributions [49]. The framework is flexible enough to accommodate any number of path selection strategies and priors. One could even stretch it to accommodate other types of mixes (even thought it is questionable why this would be interesting). Extending this model to non-global adversaries will require some key conceptual advances, and is a topic likely to require about a PhD's worth of work. The downside is that the distributions are derived through Markov-Chain Monte Carlo (MCMC) simulations, and thus no analytic form is available. As a result there is no lower bound on the security of the systems, and the job of measuring anonymity

really becomes an empirical discipline, that requires architects to test different types of parameters against different types of adversaries to find appropriate sweet spots.

- Finally the Bayesian paradigm has been applied to long-term attacks against perfect mixes [10], following the Disclosure attack models [1]. This performs as well as any other long term attack, and is much more flexible to accommodate other models of user behaviour or system behaviour. Yet it relies on MCMC to provide a measure of anonymity for a single observation trace, and thus little intuition about lowed bounds on security.

What has become clear for this line of work, is that deriving the posterior belief of the adversary, or even the likelihood of a particular trace, from a complex system is a very hard task. It is very unlikely that there will be any systematic improvement or success, unless systems are designed from the ground up to facilitate this task [8]. In those cases MCMC techniques could be used for the experimental evaluation of designs. The constraints necessary for deriving analytic expressions seem incompatible with the complexities and systems constraints of real systems.

4 Degradation Over Time

Dogan Kestogan was one of the first to point out that repeated use of an anonymity channel, even if it is perfect will eventually leak any persistent patterns of communications [3, 27]. From there, two schools of thought developed: the first extends the original Disclosure attacks, into Hitting set attacks and their approximations [30, 31]; the second, based on statistical disclosure attacks, takes a fast and loose approach to the model (and linearises it based on expectations), to infer long-term patterns [6, 9, 32]. It turns out that the entropy based metrics do not provide any insight into how fast these attacks lead to the deanonymization of a specific message, that has to go through the inference of a persistent relationship. Intuitively, even if the entropy metrics give some idea about the relationship between a specific action and a specific actor, they are poor that extending this to a relationship between a specific action and a internal profile of persistent sending. In this particular case we have to conclude that even correctly applied the entropy metrics are simply measuring the wrong thing.

It should be a bit of an embarrassment to our community that we do not have clear, generic bounds, on how quickly we expect a relationship to be disclosed, given the repeated use of a perfect anonymity channel. We have bounds on the performance of specific attacks including the Hitting Set Attack and the vanilla Statistical Disclosure attack, but those are tied to the techniques, and do not represent a fundamental limit (attempts include [28, 37]). This gap in the literature is partly due to the fact that the measures we have used for anonymity, simply do not compose under sequential execution of the channel.

One key question that has been posed by Matt Wright and Nikita Borisov, is whether differential privacy (DP) [19] could be used as a measure of anonymity. So, we present here a first differentially private bound on the security of a perfect anonymous channel, that composes nicely to allow us to reason about long term attacks (this differs from [2] that reasons about a single round). We note that the application of DP in this context is a massive departure from orthodoxy in either communities, and its meaning will have to be scrutinised quite closely.

Theorem 1. *Consider a user Alice that sends a single message through a perfect anonymous channel along with V_o other honest users. Alice sends a message to Bob with probability $0 \leq p_{AB} < \frac{\epsilon}{1+\epsilon}$, and others send messages to Bob each independently with probability p_{OB}. For any valid values of p_{AB} and p'_{AB}, and for any number of observed messages V_B from the channel to Bob (where $0 \leq V_B \leq V_o + 1$) it holds that:*

$$\Pr[V_B|p_{AB}] \leq e^{\epsilon} \Pr[V_B|p'_{AB}] + \delta,$$

where $\delta = \frac{\epsilon \cdot e^{2p_{OB}(1+\epsilon)}}{(1+\epsilon)^2 V_o p_{OB}(1-p_{OB})}$

Proof. See appendix for full proof, and to discover that ϵ is tight for small p_{OB}, but meaningless for larger ones; and that the bound on δ is very loose indeed. □

The upper bound on p_{AB} is key to interpreting the theorem. It basically provides a bound on the likelihood ratios of observed events, when it comes to distinguishing whether Alice sends to Bob with any two rates below that known bound $0 \leq p_{AB} < \frac{\epsilon}{1+\epsilon}$. As Smith [26] shows this relates directly to a bound over the ratio of posterior distributions.

One could interpret this theorem as stating that a perfect channel provides (ϵ, δ)-differential privacy. Yet there is a caveat: we have assumed that the adversary has no side information about who others are sending messages to, besides the known probability p_{OB} they send to Bob. This restriction goes away if we modify the perfect anonymity channel in the following manner: it receives one message from Alice, and any number of messages from other users; then it generates internally V_o dummy messages, each with a probability p_{OB} of being sent to Bob. Since now the "noise" is internal to the security mechanism, any side information (such as who others are sending to) cannot violate the given bound. We can further modify the protocol to ensure that Alice sends to Bob with at most probability $\frac{\epsilon}{1+\epsilon}$ through the use of source cover traffic. Thus, if we assume in the spirit of differential private mechanism that the adversary has arbitrary side information the above theorem guide us as to how much cover traffic is needed to achieve a lever of privacy equivalent to using an honest mix.

The standard differential privacy composition property applies. A number k of repeated communications through the channel compose to provide a $(k \cdot \epsilon, k \cdot \delta)$ differentially private mechanism. It is interesting to note that the ϵ depends only on the range of probabilities p_{AB}, and only δ depends on the security parameter of the mix, or the sending probability of other users. A quick peak in the proof illuminates this apparent paradox: a lot of information leaks in case $V_B = 0$, which is likely in common configurations of anonymity systems (a lot of information also leaks when $V_B = V_o + 1$ is maximized, but that is extremely unlikely). We note that the bound that yields δ is very loose, and could be seriously improved.

5 Onion Routing

Onion routing [47] has so far presented a unique challenge when it comes to measuring the quality of protection it offers, as aptly described in Syverson's paper "Why I'm not an Entropist" [46]. While there are many reasons (see above) to consider entropy metrics inadequate, we believe the reasons they cannot be easily applied to Onion Routing seems to be as much a weakness in Onion Routing as the poverty of the metrics themselves.

From the very start Onion Routing security has been defined by what it is *not* secure against: if an adversary is able to "observe" the start and the end of a circuit, then they can link them together and trivially infer who is communicating with whom [48]. Thus the security of onion routing is based on making this unlikely, either through distributing entry and exit points [20] or doing path selection [25] using trust metrics. Yet, those counter measures cannot protect, with certainty, against even a trivial attacker: If Alice wants to browse Bob's website, she is absolutely vulnerable to Charlie that happens to observer *only* her internet connection and Bob's internet connection. In this case the onion routing system can do *nothing* to protect the communication. Given this, it is clear that conventional measures of security that attempt to hide actions using a mechanism tuned by some security parameter are just not applicable, making even the definition, let alone the measurement, of the security of onion routing difficult.

Besides the above sticking point, a plethora of papers have studied the susceptibility of onion routing to different entities having some access to large fractions of input-output links, through AS topology [20] or sampling flows in the core of the network [35]. Tragedy turns into farce, since even very distorted observations such as indirect measurements of load [34], temperature-induced clock skew [33], or remote DoS attacks on network routers [23] can in fact induce "observations", and be vectors for attack. In fact the adversary learning any function of the organic or induced load on the two ends of an onion routing link will eventually lead to some compromises.

Now, besides the above fundamental limit the internal engineering of current onion routing sys-

tems [17] in itself may facilitate attacks. For example Tor currently uses three semi-stable mid-term "guards" for each client to mitigate some of the above attacks. These three guards act as an internal fingerprint for the client: if an adversary has discovered a client's guards (by observing their hotel room connection for example), they can re-identify that client if they assume a different network address (for example at home). These semi-stable guards therefore make partial tracing attacks, where an adversary only discovers the guards of a node, very dangerous. Other vectors of attacks include denial of service, the selection of paths, and the scheduling of circuits.

In brief, one of the key reasons that measuring the anonymity of Tor is hard is that under many, very common, circumstances it is just not secure. Furthermore the cases under which it is secure are conditioned not on security parameters, but instead on the inherent variability of web usage patterns, physical assumptions about network leakage, web content dynamics, or the adversary not concentrating resources around targets. These are outside the control of the security designer, and should traditionally not be considered strong sources of uncertainty or replied upon for security.

6 An Inescapable Fact

Syverson is right that entropy, as a summary of the inferences of an attacker has serious shortcomings. But this must not be extended to a deeper argument about not quantifying anonymity systems on the basis of probabilistic models. In fact designers of anonymity systems must accept an inescapable fact: to reason about the security of an anonymity system *it must be possible and efficient to compute the likelihood function* of the adversary observation given the secrets and other security parameters of the system. All metrics require this: entropy, min-entropy, probability of error, bayesian inference based analysis, and differential privacy.

In many cases computing this likelihood function is hard due to incidental noise in the system; such as the exact timing of messages, the exact timing of the co-occurrence of correlated events, etc. In such cases it is prudent to analyze systems under the assumption that the adversary can trivially recover any unprotected information. For example, since Tor does not protect against timing correlations, one should assume they they always work instantly (as they may). Similarly, since it is impossible to exclude the predictable use of multiple Tor circuits to access the same resource, the security of the system under such known patterns of access must be assessed.

Finally, as it is argued in [8], it is unlikely that an arbitrary anonymous channel construction will naturally have an easy to extract likelihood function. Therefore I am looking forward to the next 10 years of engineering involving collaborations to co-design efficient channels that are also easy to analyse from the ground up.

References

[1] Dakshi Agrawal and Dogan Kesdogan. Measuring anonymity: The disclosure attack. *IEEE Security & Privacy*, 1(6):27–34, 2003.

[2] Michael Backes, Aniket Kate, Praveen Manoharan, Sebastian Meiser, and Esfandiar Mohammadi. Anoa: A framework for analyzing anonymous communication protocols. 2013.

[3] Oliver Berthold, Andreas Pfitzmann, and Ronny Standtke. The disadvantages of free mix routes and how to overcome them. In Federrath [21], pages 30–45.

[4] Konstantinos Chatzikokolakis, Tom Chothia, and Apratim Guha. Statistical measurement of information leakage. In Javier Esparza and Rupak Majumdar, editors, *TACAS*, volume 6015 of *Lecture Notes in Computer Science*, pages 390–404. Springer, 2010.

[5] Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Prakash Panangaden. Probability of error in information-hiding protocols. In *CSF*, pages 341–354. IEEE Computer Society, 2007.

[6] George Danezis. Statistical disclosure attacks. In Gritzalis et al. [24], pages 421–426.

[7] George Danezis, Claudia Díaz, Emilia Käsper, and Carmela Troncoso. The wisdom of crowds: Attacks and optimal constructions. In Michael Backes and Peng Ning, editors, *ESORICS*, volume 5789 of *Lecture Notes in Computer Science*, pages 406–423. Springer, 2009.

[8] George Danezis and Emilia Käsper. The dangers of composing anonymous channels. In Matthias Kirchner and Dipak Ghosal, editors, *Information Hiding*, volume 7692 of *Lecture Notes in Computer Science*, pages 191–206. Springer, 2012.

[9] George Danezis and Andrei Serjantov. Statistical disclosure or intersection attacks on anonymity systems. In Fridrich [22], pages 293–308.

[10] George Danezis and Carmela Troncoso. Vida: How to use bayesian inference to de-anonymize persistent communications. In Ian Goldberg and Mikhail J. Atallah, editors, *Privacy Enhancing Technologies*, volume 5672 of *Lecture Notes in Computer Science*, pages 56–72. Springer, 2009.

[11] Yves Deswarte, Frédéric Cuppens, Sushil Jajodia, and Lingyu Wang, editors. *Information Security Management, Education and Privacy, IFIP 18th World Computer Congress, TC11 19th International Information Security Workshops, 22-27 August 2004, Toulouse, France*. Kluwer, 2004.

[12] Claudia Díaz and Bart Preneel. Reasoning about the anonymity provided by pool mixes that generate dummy traffic. In Fridrich [22], pages 309–325.

[13] Claudia Díaz and Bart Preneel. Taxonomy of mixes and dummy traffic. In Deswarte et al. [11], pages 215–230.

[14] Claudia Díaz and Andrei Serjantov. Generalising mixes. In Dingledine [16], pages 18–31.

[15] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In Dingledine and Syverson [18], pages 54–68.

[16] Roger Dingledine, editor. *Privacy Enhancing Technologies, Third International Workshop, PET 2003, Dresden, Germany, March 26-28, 2003, Revised Papers*, volume 2760 of *Lecture Notes in Computer Science*. Springer, 2003.

[17] Roger Dingledine, Nick Mathewson, and Paul F. Syverson. Tor: The second-generation onion router. In *USENIX Security Symposium*, pages 303–320. USENIX, 2004.

[18] Roger Dingledine and Paul F. Syverson, editors. *Privacy Enhancing Technologies, Second International Workshop, PET 2002, San Francisco, CA, USA, April 14-15, 2002, Revised Papers*, volume 2482 of *Lecture Notes in Computer Science*. Springer, 2003.

[19] Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *ICALP (2)*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2006.

[20] Nick Feamster and Roger Dingledine. Location diversity in anonymity networks. In Vijay Atluri, Paul F. Syverson, and Sabrina De Capitani di Vimercati, editors, *WPES*, pages 66–76. ACM, 2004.

[21] Hannes Federrath, editor. *Designing Privacy Enhancing Technologies, International Workshop on Design Issues in Anonymity and Unobservability, Berkeley, CA, USA, July 25-26, 2000, Proceedings*, volume 2009 of *Lecture Notes in Computer Science*. Springer, 2001.

[22] Jessica J. Fridrich, editor. *Information Hiding, 6th International Workshop, IH 2004, Toronto, Canada, May 23-25, 2004, Revised Selected Papers*, volume 3200 of *Lecture Notes in Computer Science*. Springer, 2004.

[23] Xun Gong, Nikita Borisov, Negar Kiyavash, and Nabil Schear. Website detection using remote

traffic analysis. In Simone Fischer-Hübner and Matthew Wright, editors, *Privacy Enhancing Technologies*, volume 7384 of *Lecture Notes in Computer Science*, pages 58–78. Springer, 2012.

[24] Dimitris Gritzalis, Sabrina De Capitani di Vimercati, Pierangela Samarati, and Sokratis K. Katsikas, editors. *Security and Privacy in the Age of Uncertainty, IFIP TC11 18th International Conference on Information Security (SEC2003), May 26-28, 2003, Athens, Greece*, volume 250 of *IFIP Conference Proceedings*. Kluwer, 2003.

[25] Aaron Johnson, Paul F. Syverson, Roger Dingledine, and Nick Mathewson. Trust-based anonymous communication: adversary models and routing algorithms. In Yan Chen, George Danezis, and Vitaly Shmatikov, editors, *ACM Conference on Computer and Communications Security*, pages 175–186. ACM, 2011.

[26] Shiva Prasad Kasiviswanathan and Adam Smith. A note on differential privacy: Defining resistance to arbitrary side information. *CoRR*, abs/0803.3946, 2008.

[27] Dogan Kesdogan, Dakshi Agrawal, and Stefan Penz. Limits of anonymity in open environments. In Petitcolas [38], pages 53–69.

[28] Dogan Kesdogan, Dakshi Agrawal, Dang Vinh Pham, and Dieter Rautenbach. Fundamental limits on the anonymity provided by the mix technique. In *IEEE Symposium on Security and Privacy*, pages 86–99. IEEE Computer Society, 2006.

[29] Dogan Kesdogan, Jan Egner, and Roland Büschkes. Stop-and-go-mixes providing probabilistic anonymity in an open system. In David Aucsmith, editor, *Information Hiding*, volume 1525 of *Lecture Notes in Computer Science*, pages 83–98. Springer, 1998.

[30] Dogan Kesdogan, Daniel Mölle, Stefan Richter, and Peter Rossmanith. Breaking anonymity by learning a unique minimum hitting set. In Anna E. Frid, Andrey Morozov, Andrey Rybalchenko, and Klaus W. Wagner, editors, *CSR*, volume 5675 of *Lecture Notes in Computer Science*, pages 299–309. Springer, 2009.

[31] Dogan Kesdogan and Lexi Pimenidis. The hitting set attack on anonymity protocols. In Fridrich [22], pages 326–339.

[32] Nick Mathewson and Roger Dingledine. Practical traffic analysis: Extending and resisting statistical disclosure. In David Martin and Andrei Serjantov, editors, *Privacy Enhancing Technologies*, volume 3424 of *Lecture Notes in Computer Science*, pages 17–34. Springer, 2004.

[33] Steven J. Murdoch. Hot or not: revealing hidden services by their clock skew. In Ari Juels, Rebecca N. Wright, and Sabrina De Capitani di Vimercati, editors, *ACM Conference on Computer and Communications Security*, pages 27–36. ACM, 2006.

[34] Steven J. Murdoch and George Danezis. Low-cost traffic analysis of tor. In *IEEE Symposium on Security and Privacy*, pages 183–195. IEEE Computer Society, 2005.

[35] Steven J. Murdoch and Piotr Zielinski. Sampled traffic analysis by internet-exchange-level adversaries. In Nikita Borisov and Philippe Golle, editors, *Privacy Enhancing Technologies*, volume 4776 of *Lecture Notes in Computer Science*, pages 167–183. Springer, 2007.

[36] Luke O'Connor. On blending attacks for mixes with memory. In Mauro Barni, Jordi Herrera-Joancomartí, Stefan Katzenbeisser, and Fernando Pérez-González, editors, *Information Hiding*, volume 3727 of *Lecture Notes in Computer Science*, pages 39–52. Springer, 2005.

[37] Luke O'Connor. Entropy bounds for traffic confirmation. *IACR Cryptology ePrint Archive*, 2008:365, 2008.

[38] Fabien A. P. Petitcolas, editor. *Information Hiding, 5th International Workshop, IH 2002, Noordwijkerhout, The Netherlands, October 7-9,*

2002, Revised Papers, volume 2578 of *Lecture Notes in Computer Science*. Springer, 2003.

[39] Andreas Pfitzmann and Marit Köhntopp. Anonymity, unobservability, and pseudonymity - a proposal for terminology. In Federrath [21], pages 1–9.

[40] Michael K. Reiter and Aviel D. Rubin. Anonymous web transactions with crowds. *Commun. ACM*, 42(2):32–38, 1999.

[41] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In Dingledine and Syverson [18], pages 41–53.

[42] Andrei Serjantov, Roger Dingledine, and Paul F. Syverson. From a trickle to a flood: Active attacks on several mix types. In Petitcolas [38], pages 36–52.

[43] Andrei Serjantov and Richard E. Newman. On the anonymity of timed pool mixes. In Gritzalis et al. [24], pages 427–434.

[44] Vitaly Shmatikov and Ming-Hsiu Wang. Measuring relationship anonymity in mix networks. In Ari Juels and Marianne Winslett, editors, *WPES*, pages 59–62. ACM, 2006.

[45] Sandra Steinbrecher and Stefan Köpsell. Modelling unlinkability. In Dingledine [16], pages 32–47.

[46] Paul Syverson. Why im not an entropist. In *Security Protocols XVII*, pages 213–230. Springer, 2013.

[47] Paul F. Syverson, David M. Goldschlag, and Michael G. Reed. Anonymous connections and onion routing. In *IEEE Symposium on Security and Privacy*, pages 44–54. IEEE Computer Society, 1997.

[48] Paul F. Syverson, Gene Tsudik, Michael G. Reed, and Carl E. Landwehr. Towards an analysis of onion routing security. In Federrath [21], pages 96–114.

[49] Carmela Troncoso and George Danezis. The bayesian traffic analysis of mix networks. In Ehab Al-Shaer, Somesh Jha, and Angelos D. Keromytis, editors, *ACM Conference on Computer and Communications Security*, pages 369–379. ACM, 2009.

[50] Matthew Wright, Micah Adler, Brian Neil Levine, and Clay Shields. An analysis of the degradation of anonymous protocols. In *NDSS*. The Internet Society, 2002.

A Proof of theorems

Theorem 1. *Consider a user Alice that sends a single message through a perfect anonymous channel along with V_o other honest users. Alice sends a message to Bob with probability $0 \leq p_{AB} < \frac{\epsilon}{1+\epsilon}$, and others send messages to Bob each independently with probability p_{OB}. For any valid values of p_{AB} and p'_{AB}, and for any number of observed messages V_B from the channel to Bob (where $0 \leq V_B \leq V_o + 1$) it holds that:*

$$\Pr[V_B|p_{AB}] \leq e^{\epsilon} \Pr[V_B|p'_{AB}] + \delta,$$

where $\delta = \frac{\epsilon \cdot e^{2p_{OB}(1+\epsilon)}}{(1+\epsilon)^2 V_o p_{OB}(1-p_{OB})}$

Proof. Consider the distribution over the number of messages destined to Bob resulting from a single use of the perfect anonymity system. It is the result of the sum of the potential message from Alice, and the remaining messages from others. Both are random variables and we can derive the distribution of the sum through the convolution of the distribution of the separate random variables:

$$\Pr[V_B|p_{AB}, \ldots] = \sum_{V_{AB}=\max(0,V_B-V_o)}^{\min(1,V_B)} p_{AB}^{V_{AB}} \bar{p}_{AB}^{(1-V_{AB})} \binom{V_o}{V_B - V_{AB}} p_{OB}^{(V_B-V_{AB})} \bar{p}_{OB}^{(V_o-V_B+V_{AB})} \tag{1}$$

where $\bar{p}_{AB} = 1 - p_{AB}$ and $\bar{p}_{OB} = 1 - p_{OB}$.

We want to determine ϵ and δ an such that $\Pr[V_B|p_{AB}] \leq e^{\epsilon} \Pr[V_B|p'_{AB}] + \delta$ for all $0 \leq V_B \leq V_o + 1$ and for any two possible p_{AB}, p'_{AB} subject to the constraint $p_{AB}, p'_{AB} < \frac{\epsilon}{1+\epsilon}$. We will do this by considering the cases where $V_B = 0$, $0 < V_B < V_o + 1$, and $V_B = V_o + 1$ separately.

In the case $V_B = 0$, it must be that $V_{AB} = 0$ (reducing the sum to a single element independent of p_{AB}) and the ratio becomes:

$$\frac{\Pr[V_B = 0|p_{AB}, \ldots]}{\Pr[V_B = 0|p'_{AB}, \ldots]} = \frac{\bar{p}_{AB}}{\bar{p}'_{AB}} \leq \frac{1}{\bar{p}'_{AB}} = \frac{1}{1 - \frac{\epsilon}{1+\epsilon}} = 1 + \epsilon \leq e^{\epsilon} \tag{2}$$

In case $0 < V_B < V_o + 1$ we note that the probability of V_B becomes:

$$\Pr[V_B|p_{AB}, \ldots] = \binom{V_o}{V_B} p_{OB}^{V_B} \bar{p}_{OB}^{(V_o-V_B)} \left[1 + \left(\frac{V_B}{V_o - V_B + 1} \frac{\bar{p}_{OB}}{p_{OB}} - 1\right) p_{AB}\right] \tag{3}$$

We define x as $x = \frac{V_B}{V_o - V_B + 1} \frac{\bar{p}_{OB}}{p_{OB}}$, and split two cases, when $x - 1 \leq 0$ and when $x - 1 > 0$.

When $x - 1 < 0$, it implies that $V_B > (V_o + 1)p_{OB}$, and we can bound the ratio of probabilities for different p_{AB}, p'_{AB}. We note that this ratio is minimized when $V_B = 1$, and substitute it into x:

$$\frac{\Pr[V_B|p_{AB}, \ldots]}{\Pr[V_B|p'_{AB}, \ldots]} \leq \frac{1}{1 + (x-1)p_{AB}} = \frac{V_o}{V_o - V_o p_{AB} + \left(\frac{1}{p_{OB}} - 1\right) p_{AB}} \leq \frac{V_o}{V_o - V_o \cdot p_{AB}} \tag{4}$$

$$= 1 + e \leq e^{\epsilon} \tag{5}$$

When $x - 1 \geq 0$, it implies that $V_B < (V_o + 1)p_{OB}$ and things get tricky. We can show that if $V_B < \frac{p_{OB}(V_o+1)(2+\epsilon)}{1+p_{OB}(1+\epsilon)}$ the ratio of likelihoods can be bound as:

$$\frac{\Pr[V_B|p_{AB}, \ldots]}{\Pr[V_B|p'_{AB}, \ldots]} \leq 1 + (x-1)p_{AB} \leq 1 + \epsilon < e^{\epsilon} \tag{6}$$

because

$$1 + (1 - x)p_{AB} \leq 1 + \epsilon \Leftrightarrow x \leq 1 + \epsilon \Leftrightarrow V_B < \frac{p_{OB}(V_o + 1)(2 + \epsilon)}{1 + p_{OB}(1 + \epsilon)} \tag{7}$$

This bound does not hold if V_B is larger, and thus we need to bound the statistical difference in this case:

$$\Pr[V_B|p_{AB}, \ldots] - e^{\epsilon} \Pr[V_B|p'_{AB}, \ldots] \leq \Pr[V_B|p_{AB}, \ldots] - \Pr[V_B|p'_{AB}, \ldots] = \tag{8}$$

$$= \binom{V_o}{V_B} p_{OB}^{V_B} \bar{p}_{OB}^{(V_o - V_B)} (x - 1)(p_{AB} - p'_{AB}) \tag{9}$$

$$\leq \binom{V_o}{V_B} p_{OB}^{V_B} \bar{p}_{OB}^{(V_o - V_B)} (x - 1)(\frac{\epsilon}{1 + \epsilon}) \tag{10}$$

$$\leq \binom{V_o}{V_B} p_{OB}^{V_B} \bar{p}_{OB}^{(V_o - V_B)} \cdot \epsilon \tag{11}$$

Since this is maximized when $V_B = \frac{p_{OB}(V_o+1)(2+\epsilon)}{1+p_{OB}(1+\epsilon)}$, at which point $x - 1 = 1 + \epsilon$. At this point we could just plug this V_B into the expression above to get an accurate value for δ. Instead we will opt for having an "elegant" theorem, and bound this expression using Chebyshev bounds on the tails of the binomial distribution.

We note that the binomial distribution could be bound by its tail:

$$\binom{V_o}{V_B} p_{OB}^{V_B} \bar{p}_{OB}^{(V_o - V_B)} \cdot \epsilon = \Pr[V_B = a|V_o, p_{OB}] \cdot \epsilon \leq \Pr[V_B > b|V_o, p_{OB}] \cdot \epsilon \tag{12}$$

if $a \in (b, V_o)$. We define $b = \frac{p_{OB}V_o(2+\epsilon)}{1+p_{OB}(1+\epsilon)} < \frac{p_{OB}(V_o+1)(2+\epsilon)}{1+p_{OB}(1+\epsilon)}$. The Chebyshev bound on the tails of the binomial distribution only depends on its mean $\mu = V_o p_{OB}$ and variance $\sigma^2 = V_o p_{OB} \bar{p}_{OB}$.

$$\Pr[V_B > b] \cdot \epsilon \leq \frac{\sigma^2}{(\mu - b)^2} \cdot \epsilon \tag{13}$$

$$= \frac{\epsilon}{V_o p_{OB} \bar{p}_{OB}} \left[\frac{1 + p_{OB}(1 + \epsilon}{1 + \epsilon}\right]^2 \tag{14}$$

$$\leq \frac{\epsilon \cdot e^{2p_{OB}(1+\epsilon)}}{V_o p_{OB} \bar{p}_{OB}(1 + \epsilon)^2} \tag{15}$$

This becomes our value for δ.

The very final case of the proof concerns $V_B = V_o + 1$. The probability of this happening is small, therefore:

$$\Pr[V_B|p_{AB}, \ldots] - e^{\epsilon} \Pr[V_B|p'_{AB}, \ldots] \leq \Pr[V_B|p_{AB}, \ldots] - \Pr[V_B|p'_{AB}, \ldots] \tag{16}$$

$$= (p_{AB} - p'_{AB}) p_{OB}^{V_o} \tag{17}$$

$$\leq \frac{\epsilon}{1 + \epsilon} p_{OB}^{V_o} < \delta \tag{18}$$

This concludes the proof since in all cases either we have bound the ratio of likelihoods by e^{ϵ} or we have shown that the differences of probabilities are less than δ.

□