The Byzantine Postman Problem

Len Sassaman Bart Preneel
K.U. Leuven ESAT-COSIC and IBBT
Kasteelpark Arenberg 10, B-3001
Leuven-Heverlee, Belgium
len.sassaman@esat.kuleuven.be bart.preneel@esat.kuleuven.be

Abstract

Over the last several decades, there have been numerous proposals for systems which can preserve the anonymity of the recipient of some data. Some have involved trusted third-parties or trusted hardware; others have been constructed on top of link-layer anonymity systems or mix-nets.

In this paper, we evaluate a pseudonymous message system which takes the different approach of using Private Information Retrieval (PIR) as its basis. We expose a flaw in the system as presented: it fails to identify Byzantine servers. We provide suggestions on correcting the flaw, while observing the security and performance trade-offs our suggestions require.

1 Introduction

Several proposals have been made for the use of private information retrieval (PIR) [7] primitives to build secure, fault-tolerant pseudonymous mail retrieval systems [8, 3, 15, 20].

PIR-based pseudonym (or *nym*) servers have several significant advantages over nym servers based on other technologies. PIR protocols can be designed to offer *information-theoretic security*, i.e., assuming that the system is correct, an attacker with unlimited computational power cannot defeat the system merely by virtue of being able to perform calculations which reveal the private information. Other PIR protocols merely offer computational security: in Computational PIR systems [6], the privacy of the PIR query is protected only against an adversary restricted to polynomial-time computational capability. CPIR-based solutions have the significant advantage that they can be performed using a single server, and do not require distribution of trust to ensure that the information retrieval requests remain private. However, such systems presently have prohibitive computational cost on commodity hardware.

The most recent proposal for a nym server based on PIR with information-theoretic security, the Pynchon Gate [20], offers greater robustness, stronger anonymity assurances, and better traffic analysis resistance than previously proposed pseudonym systems. However, it contains a serious flaw in its protocol which can be used to launch a denial of service attack against the system, rendering it unusable. Furthermore, the attack is not merely limited to decreased utility of the system; due to the network-effects properties of anonymity systems, denying service to one set of users can effectively weaken the anonymity provided to a different set of users [1]. We identify this denial of service attack, evaluate the extent of the problem, and briefly consider solutions which may be considered as a means of making the protocol immune to the attack.

2 Background on Nym Servers

Pseudonymous messaging services allow users to send messages that originate at a pseudonymous address (or "nym") unlinked to the user, and to receive messages sent to that address, without allowing an attacker to deduce which users are associated with which pseudonyms. These systems can be used for parties to communicate without revealing their identities, or as a building-block for other systems that need a bi-directional anonymous communication channel, such as Free Haven [11].

3 Background on The Pynchon Gate

To address the reliability problems of silent node failure, as well as the serious security problems of statistical disclosure and end-to-end traffic analysis, Sassaman et al. propose a complete architectural design of a PIR-based pseudonym service offering information-theoretic protection, called the Pynchon Gate [20].

3.1 Architecture overview

The architecture of the Pynchon Gate consists of an Internet-facing component referred to as the "nym server", which receives messages addressed to users of the system and acts as a gateway between the pseudonym service and other Internet services such as email. Behind the nym server is a component known as the "collator", which structures the incoming messages in the form of a three-level hash tree, which is then replicated to a series of mutually untrusted distribution databases referred to as "distributors".

Email addressed to a specific pseudonym is stored in a specific location in the database, such that the owner of the pseudonym knows what information to request to obtain his message. Using the PIR protocol described in Section 3.2, the user submits a PIR query to ℓ distributors, and his message is returned with none of the distributors able to deduce any information about the user's query unless all ℓ distributors collude. This form of PIR is referred to as an **information-theoretic ($\ell-1$)-private ℓ-server PIR** protocol.

3.2 The Pynchon Gate PIR Protocol

The protocol runs as follows: after choosing distributors, the client establishes an encrypted connection to each (e.g., using TLS [10]). These connections must be unidirectionally authenticated to prevent man-in-the-middle attacks, and can be made sequentially or in parallel.

The client sends a different "random-looking" bit vector $\vec{\nu}_{s\beta}$ to each distributor s for each message block β to be retrieved. Each bit vector has a length equal to the number of message blocks in the database. Each distributor s then computes $R(\vec{\nu}_{s\beta})$ as the exclusive-OR of all message blocks whose positions are set to 1 in $\vec{\nu}_{s\beta}$. The resulting value is then returned to the client.

Thus, in order to retrieve the β'th message block, the client need only choose the values of $\vec{\nu}_{s\beta}$ such that when all $\vec{\nu}_{s\beta}$ are XORed together, all values are 0 at every position except β. (For security, $\ell - 1$ of the vectors should be generated randomly, and the bit vectors should be sent in a random order so that the ℓ'th, specially-crafted vector cannot be distinguished.) When the client receives the corresponding $R(\vec{\nu}_{s\beta})$ values, she can XOR them to compute the message block's contents.

3.3 Byzantine Server Protection

In a distributed-trust anonymity system such as the Pynchon Gate, there exists the possibility that some servers may be *Byzantine*, i.e., they may behave incorrectly, either due to intentional malice or simple error.[1] In the case of the Pynchon Gate, the Byzantine behavior we are concerned with is an incorrect response to a PIR query of a distributor's database.

All n distributors in the system have the exact same copy of the database, and the system is designed such that any attempt by a Byzantine server to modify its response to the PIR query will be detected by the user when he verifies the root of the hash tree. This is crucial to preserving the anonymity properties of the system, for if an attacker can alter a message or observe the cleartext of a message, he may be able to later link an input message with a given output retrieved by the nym holder.

The Pynchon Gate's message and link encryption prevents an attacker from observing the cleartext of a message. Active attacks that are dependent upon the attacker's ability to alter some of the data being transmitted to the user such that the attacker may later link the user to his pseudonym based either on a variance in the user's response to altered versus unaltered data, or by simply recognizing the product of the altered data as it is processed by the system (collectively known as *tagging attacks* [13]), are ineffective, as TLS protects data integrity on the wire. Thus, any tagging attacks an attacker wished to attempt against a user would have to occur through the use of a corrupt distributor. To protect against the case where a distributor provides (intentionally or otherwise) an incorrect response to the PIR query, the client verifies that the hash of the message block it has received can be authenticated through the hash tree with the verified hash root.

3.4 A Remaining Byzantine Server Attack

We present the following attack not prevented by the hash tree verification system: a corrupt distributor can, through malice or error, create a denial of service attack on the system by responding with incorrect data to a client's query. While the client will detect that the message block is invalid after performing the final step of the PIR protocol in Subsection 3.2, and thus can conclude that *some* server was Byzantine, the client cannot determine *which* server or servers returned the incorrect response. The client cannot safely pass the message block contents (assuming they consist of anything other than garbage) to the user, lest the user's anonymity be potentially compromised.

Furthermore, if attacks on portions of the pseudonymity infrastructure affect some users differently than others, an attacker may exploit such attacks on components of the system to facilitate an intersection attack against a user of the system as a whole [12]. In the Pynchon Gate, if a Byzantine distributor selectively performed denial of service attacks against certain users by returning garbage results to their queries, but correctly responded to other users' queries, the attacker would increase his chances of learning the identity of certain users, based on which users responded to messages that were successfully delivered.[2] In other cases, a passive adversary could observe the actions of Byzantine servers not under his control (and perhaps not even behaving maliciously, but simply incorrectly) to help facilitate intersection attacks [23]. Additionally, if a user cannot know with confidence which server is behaving in a Byzantine fashion, she is more likely to change the nodes she uses on a regular basis, both increasing her exposure to long-term intersection attacks and increasing the probability of selecting a server-set that consists of nodes operated entirely by a single adversary.

[1] This concern is present in many other anonymity systems, including Chaumian mix-nets [5, 18, 9] and systems built on top of them [17, 16].

[2] This type of attack is present (in a slightly different form) in non-PIR-based nym server systems as well. For instance, in a reply-block system, an attacker could disable certain mixes and observe which nyms ceased receiving traffic. If the nym holder has a fixed-route reply-block, this would enable the attacker to identify the mixes used in the nym holder's reply-block path, and increase his chances of successfully linking the nym with the nym holder's true name [22].

4 Byzantine Server Detection

Ideally, there would exist a way to identify an individual Byzantine server without modifying the existing threat model or positive security properties of the Pynchon Gate. This is a challenging problem to solve with the existing XOR-based PIR protocol, which makes verifying the results of a PIR query returned by a specific distributor impossible. (The client does not know what a "correct" response $R(\vec{\nu}_{s\beta})$ from any given distributor should look like; only that

$$R(\vec{\nu}_{s_1\beta}) \oplus R(\vec{\nu}_{s_2\beta}) \oplus \cdots \oplus R(\vec{\nu}_{s_\ell\beta}) = \beta' th \text{ message block}$$

and thus cannot identify which of the responses were invalid.)

4.1 Checksums on each message block

Applying traditional hashes or checksums to each message block is not a viable approach, for it is not the message blocks themselves, but the XOR of all the blocks requested from a given distributor, that is returned by that distributor.

If there exists a commitment verification function g such that $g(f(A), f(B)) = g'(A \oplus B)$ (and g can take an arbitrary number of arguments, and g' is predictable based on g), it may be possible for the collator (already trusted with the creation and signing of the hash root) to perform the commitment f on each block, and publish that value. When encountering a corrupt message block, the client could obtain all f's corresponding to the 1's in the bit vectors it sent to the distributors, calculate $g(f(A), f(B), \cdots, f(n))$ for each bit vector sent to each distributor in turn, and identify which distributor was Byzantine by observing which calculation of g did not match the corresponding calculation of g'.

We know of no such function, nor do we know if such a function would increase the cost of operating the Pynchon Gate system prohibitively, either through excess computation, bandwidth, or storage.

4.2 Alternative PIR schemes

There exist PIR schemes that incorporate Byzantine recovery as part of the protocol, such as the scheme presented by Beimel and Stahl [2], or the recent work by Goldberg [14]. Such protocols could theoretically be used in lieu of the XOR-based scheme in the Pynchon Gate and the other PIR-based pseudonym systems referenced. These protocols have the additional property of *Byzantine recovery*, where a user can still reconstruct a message block from the responses he has received, as long as some threshold of servers are not Byzantine.

These alternate protocols we have considered are k-out-of-ℓ polynomial interpolation based schemes, and therefore have a significant drawback in that the security offered by the Pynchon Gate must be weakened. In these schemes, the threshold of nodes which must collude to break the security of the system decreases compared to the simple XOR scheme in the Pynchon Gate. In a k-out-of-ℓ scheme, if *any* k servers collude, the privacy of the user is lost. As the difference of $\ell - k$ must be at least 1 in order to provide any Byzantine robustness, the best assurance the system can offer the user is protection as long as two of the distributors are honest (and this is in the weakest configuration for Byzantine robustness!) Therefore, the threat of Byzantine servers must be weighed against the probability that an adversary may control a large enough coalition of servers to satisfy k in a polynomial interpolation based PIR scheme, and the protocol parameters chosen accordingly.

Furthermore, these, like most PIR protocols, have only been evaluated for security and privacy-preserving properties. Additional considerations apply when selecting a primitive for use in an anonymity system; such considerations may not have been part of the design criteria for these protocols. Before implementing an alternate PIR-scheme as the basis for a pseudonymity service, one must consider possible attacks on the protocol which are only of concern when it is used for anonymity purposes.

4.3 A "Cut-and-Choose" solution

At a cost of increased bandwidth, a solution based on the "cut-and-choose" [19, 4] problem could be implemented. This is an appealing avenue of research, in that it allows for the introduction of a modular solution to the denial of service attack problem which ideally has no effect upon the security of the protocol. We have proposed a protocol based on this principle in a technical report [21]. The bandwidth costs could potentially be prohibitive in some circumstances with the existing protocol, however, and implementation details remain unproven.

5 Conclusions and Future Work

We have evaluated the security of the Pynchon Gate, a PIR-based pseudonymous message system, and identified a weakness in its protocol which prevents users from identifying Byzantine servers. We have described how this limitation in the system can lead to a denial of service attack or potentially be used to compromise the anonymity of the system's users.

We have offered suggestions on potential solutions to the problems in the existing system; however, we have not provided a known solution which maintains the other security properties of the original scheme and can operate as efficiently as the original scheme. Additional work must be done on the development of Private Information Retrieval protocols as anonymity primitives.

6 Acknowledgements

We would like to thank Brian Warner and Bryce Wilcox-O'Hearn for their helpful early discussions of the Byzantine Postman Problem; Elena Andreeva, David Chaum, Markulf Kohlweiss and Brian Warner for insightful comments on specialized commitment functions; Ian Goldberg for suggesting the potential of polynomial interpolation-based PIR solutions; Elena Andreeva, Ian Goldberg and David Molnar for assistance in searching the literature for pertinent work; and David Chaum, Bram Cohen and Meredith L. Patterson for inspirational discussion of the future work directions.

We would like to thank Nikita Borisov and Bram Cohen for their comments on an earlier draft of this work which significantly improved its presentation, and Meredith L. Patterson for proofreading and bibliography assistance with the final version.

This work was supported in part by the Concerted Research Action (GOA) Ambiorics 2005/11 of the Flemish Government, by the IBBT (Flemish Government) and by the IAP Programme P6/26 BCRYPT of the Belgian State (Belgian Science Policy). Additional support was provided by the EU within the PRIME Project under contract IST-2002-507591.

References

[1] Alessandro Acquisti, Roger Dingledine, and Paul Syverson. On the Economics of Anonymity. In Rebecca N. Wright, editor, *Financial Cryptography.* Springer-Verlag, LNCS 2742, 2003.

[2] A. Beimel and Y. Stahl. Robust information-theoretic private information retrieval. In S. Cimato C. Galdi G. Persiano, editor, *3rd Conf. on Security in Communication Networks*, volume 2576 of *Lecture Notes in Computer Science*, pages 326–341. Springer-Verlag, 2002.

[3] Oliver Berthold, Sebastian Clauß, Stefan Köpsell, and Andreas Pfitzmann. Efficiency improvements of the private message service. In Ira S. Moskowitz, editor, *Proceedings of Information Hiding Workshop (IH 2001)*, pages 112–125. Springer-Verlag, LNCS 2137, April 2001.

[4] Gilles Brassard, David Chaum, and Claude Cr&#233;peau. Minimum disclosure proofs of knowledge. *Journal of Computer and System Sciences*, 37(2):156–189, 1988.

[5] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 4(2), February 1981.

[6] Benny Chor and Niv Gilboa. Computationally private information retrieval (extended abstract). In *STOC '97: Proceedings of the twenty-ninth annual ACM symposium on Theory of Computing*, pages 304–313, New York, NY, USA, 1997. ACM Press.

[7] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. In *IEEE Symposium on Foundations of Computer Science*, pages 41–50, 1995.

[8] David A. Cooper and Kenneth P. Birman. Preserving privacy in a network of mobile computers. In *Proceedings of the 1995 IEEE Symposium on Security and Privacy*, May 1995.

[9] George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a Type III Anonymous Remailer Protocol. In *Proceedings of the 2003 IEEE Symposium on Security and Privacy*, May 2003.

[10] T. Dierks and C. Allen. The TLS Protocol. Request for Comments: 2246, January 1999.

[11] Roger Dingledine, Michael J. Freedman, and David Molnar. The Free Haven Project: Distributed anonymous storage service. In H. Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*. Springer-Verlag, LNCS 2009, July 2000.

[12] Roger Dingledine and Nick Mathewson. Anonymity loves company: Usability and the network effect. In *Proceedings of the Fifth Workshop on the Economics of Information Security (WEIS 2006)*, Cambridge, UK, June 2006.

[13] Roger Dingledine and Len Sassaman. Attacks on Anonymity Systems: Theory and Practice. In *Black Hat USA 2003 Briefings*, Las Vegas, NV, USA, July 2003.

[14] Ian Goldberg. Improving the Robustness of Private Information Retrieval. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy*, May 2007.

[15] L. Kissner, A. Oprea, M. Reiter, D. Song, and K. Yang. Private keywordbased push and pull with applications to anonymous communication. *Applied Cryptography and Network Security*, 2004.

[16] Nick Mathewson. Underhill: A proposed type 3 nymserver protocol specification, August 2004. `http://www.mixminion.net/nym-spec.txt`.

[17] David Mazières and M. Frans Kaashoek. The Design, Implementation and Operation of an Email Pseudonym Server. In *Proceedings of the 5th ACM Conference on Computer and Communications Security (CCS'98)*. ACM Press, November 1998.

[18] Ulf Möller, Lance Cottrell, Peter Palfrader, and Len Sassaman. Mixmaster Protocol — Version 2, July 2003. `http://www.abditum.com/mixmaster-spec.txt`.

[19] M. O. Rabin. Digitalized signatures. In R. Lipton and R. De Millo, editors, *Foundations of Secure Computation*, pages 155–166, New York, 1978. Academic Press.

[20] Len Sassaman, Bram Cohen, and Nick Mathewson. The Pynchon Gate: A Secure Method of Pseudonymous Mail Retrieval. In *Proceedings of the Workshop on Privacy in the Electronic Society (WPES 2005)*, Arlington, VA, USA, November 2005.

[21] Len Sassaman and Bart Preneel. Solving the Byzantine Postman Problem. Technical Report ESAT-COSIC 2007-004, K.U. Leuven, 2007.

[22] Andrei Serjantov, Roger Dingledine, and Paul Syverson. From a trickle to a flood: Active attacks on several mix types. In Fabien Petitcolas, editor, *Proceedings of Information Hiding Workshop (IH 2002)*. Springer-Verlag, LNCS 2578, October 2002.

[23] Matthew Wright, Micah Adler, Brian Neil Levine, and Clay Shields. Defending anonymous communication against passive logging attacks. In *Proceedings of the 2003 IEEE Symposium on Security and Privacy*, May 2003.