Analyse einer praktischen fundamentalen Schranke der anonymen Kommunikation*

Vinh Pham

vinh.pham@ur.de

Abstract: Existierende Systeme zur anonymen Kommunikation können bei offenen Nutzergruppen, wie sie im Internet vorherrschen, keine informationstheoretisch perfekte Anonymität gewährleisten. Daher ist es notwendig die Schranke der Anonymität in diesem Fall zu bestimmen. Diese Arbeit analysiert diese Schranke anhand eines konkreten Anonymitätssystems, welches Chaum Mix genannt wird. Ein *Mix* ist ein System zur anonymen Netzwerkkommunikation, welches in jeder Kommunikationsrunde die Sender (Subjekte) von Nachrichten in eine *Anonymitätsmenge* einbettet, um die Zuordnung zu ihren Empfängern (Attributen) zu verdecken. Anoymitätsmengen bilden die Grundlage aller Anonymitätsysteme, um die Zuordnung zu sensiblen Attributen zu verschleiern. Die fundamentale Schranke der Anonymität wird daher durch die Schwierigkeit bestimmt aus den beobachtbaren Anonymitätsmengen und Attributen die Zuordnungen eindeutig zu rekonstruieren. In Anlehnung an Shannons Unicity-Distance bestimmen wir die Anonymitätsschranke über die minimale Anzahl an Anonymitätsmengen, die beobachtet werden muss, um die Zuordnung eindeutig aufzudecken. Wir zeigen, dass diese Aufdeckung in vielen realistischen Fällen sogar effizient berechenbar ist, obwohl sie das Lösen eines NP-vollständigen Problems erfordert.

1 Einleitung

In dieser Arbeit [Pha13] geht es um die Bestimmung (bzw. Messung) der Sicherheit von Anonymitätssystemen in Kommunikationsnetzen. Dazu analysieren wir die Schranke, ab der die Anonymität eines Subjekts durch ein System nicht mehr gewährleistet werden kann, so dass eine Deanonymisierung möglich wird. Die Erforschung dieser Anonymitätsschranke und ihrer Einflussfaktoren deckt allgemeine Schwächen der Anonymitätssysteme auf und unterstützt daher die Entwicklung von Systemen mit höherer Sicherheit.

Wir bestimmen die Anonymitätsschranke anschaulich anhand eines konkreten, aber einfachen Anonymitätssystems, das Chaum Mix [Cha81] genannt wird. Der Chaum Mix ermöglicht die Anonymiserung der Kommunikationsbeziehung zwischen Sendern und Empfängern von Nachrichten in einem unsicheren Netzwerk wie dem Internet. Abbildung 1 skizziert das Funktionsprinzip eines Mixes, das einer Wahlurne ähnelt. In jeder Kommunikationsrunde sammelt der Mix von b Sendern ($b = 3$ in Abbildung 1) ein Datenpaket gleicher Größe. Diese Pakete sind an den Mix adressiert und enthalten

*Englischer Titel der Dissertation: "Towards Practical and Fundamental Limits of Anonymity Protection"

für diesen verschlüsselt jeweils die eigentliche (mit x bezeichnete) Nachricht, sowie die Empfängeradresse. Der Mix entfernt die Verschlüsselung in den b Paketen und sortiert die daraus entnommenen (mit x gekennzeichneten) Nachrichten alphabetisch. Die Nachrichten werden in derselben Runde an die entschlüsselten Empfängeradressen weitergeleitet. Dieser Vorgang randomisiert das Aussehen und die Reihenfolge der am Mix eingehenden und ausgehenden Pakete. Ein beobachtender Angreifer (*passiver Angreifer*)[1] kann durch diese Maßnahme nur die *Anonymitätsmenge* der Sender S' und der Empfänger R' beobachten, aber nicht wer in dieser Runde mit wem kommuniziert hat.

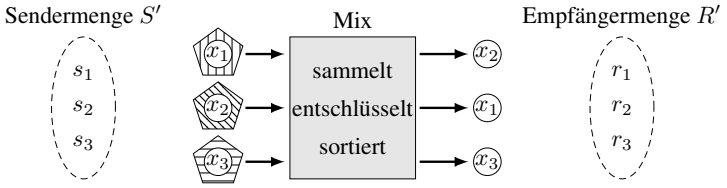


Abbildung 1: Mix Model: Jedes Muster um eine Nachricht x ist eine Verschlüsselungsebene. Mit s bzw. r wird ein Sender bzw. Empfänger bezeichnet.

Basis Modell der Anonymitätsanalyse Die Mengen (S', R') stellen die einzigen Informationen dar, die bei einer idealen Umsetzung[2] des Chaum Mixes in jeder Runde im Netzwerk beobachtbar sind. Die Anonymitätsschranke des Chaum Mixes ist daher fundamental durch die Schwierigkeit bestimmt, aus den Beobachtungen von (S', R') die Kommunikationsbeziehungen eines Senders zu identifizieren. Daher kann ohne Beeinträchtigung der Allgemeinheit die Ermittlung dieser Schranke auf die Analyse der Beobachtungen von (S', R') reduziert und vereinfacht werden. Sei *Alice* ein beliebiger Sender, dann müssen sogar nur die Mengen (S', R') betrachtet werden, in denen Alice als Sender in S' vorkommt, um ihre Kommunikationsbeziehung zu deanonymisieren. Wir nennen die Empfängermenge R' in diesem Fall eine *Observation*, so dass jede Observation wie in Abbildung 2 mindestens einen Empfänger von Alice enthält. Zur Hervorhebung wird ein Empfänger von Alice als ein *Freund* bezeichnet und mit der Variablen a adressiert, während die Variable r verwendet wird, wenn keine Unterscheidung erforderlich ist.

Erweitertes Modell der Anonymitätsanalyse Das Basis Modell ist der initiale Ansatz für die Anonymitätsanalyse in dieser Arbeit. Wir erweitern dieses Modell um die Möglichkeit von *fehlerhaften Observationen*, welche entgegen der Erwartung des Angreifers keine Freunde enthalten. Fehlerhafte Observation können entstehen, wenn z.B. der Angreifer zu schwach ist, um die Anonymitätsmengen (S', R') vollständig zu beobachten. Des Weiteren können sie induziert werden, wenn das Kommunikationsprotokoll

[1] Wir betrachten einen passiven Angreifer, weil seine Angriffe schwer erkennbar und schwer vermeidbar sind. Angriffe von aktiven Angreifer können hingegen durch die CUVE [KP06] Anforderung des Chaum Mixes erkannt werden.

[2] Diese schließt Schwächen aus, die auf die Implementierung, oder z.B. kryptographische Protokolle zurückzuführen sind, um ausschließlich die Anonymitätsfunktion zu bewerten.

Scheinnachrichten vorsieht, oder Mix Varianten wie z.B. Mixmaster [DDM03] eingesetzt werden, die Nachrichten indeterministisch weiterleiten[3].

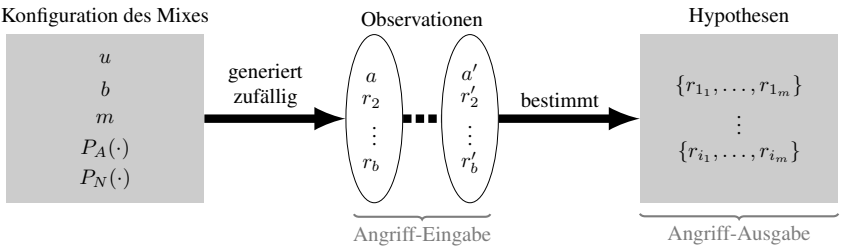


Abbildung 2: Anonymitätsanalyse: Mit r bzw. a wird ein Empfänger bzw. ein Freund bezeichnet. Aus Observationen werden Hypothesen für die möglichen m Freunde von Alice gebildet.

Wahl des Modells Die Arbeit betrachtet den Chaum Mix, weil es ein einfaches Anonymitätssystem darstellt. Dadurch sind die Einflussfaktoren auf die Anonymitätsmengen (S', R') übersichtlich. Dieses vereinfacht die Analyse der Sicherheit für unterschiedliche Einstellungen der Anonymitätsmengen. Zugleich ist der Chaum Mix die theoretische Grundlage vieler anonymer Kommunikationssysteme [EY09, DD08], so dass auch eine praktische Anwendung der Forschungsergebnisse begünstigt wird.

Erwartungswert der Anonymitätsschranke Abbildung 2 modelliert über die *Konfiguration des Mixes* die Einflussfaktoren auf die Anonymität. Damit studieren wir die erwartete Wirkung beliebiger Konfigurationen auf die Anonymitätsschranke, um den Entwurf sichererer Systeme zu unterstützen. Die Konfiguration des Mixes enthält die Gesamtzahl aller möglichen Empfänger u im System (d.h. $u = |\{r \mid r \text{ kann Nachrichten empfangen}\}|$), die Größe der Anonymitätsmenge b und die Anzahl der Freunde von Alice m. Des Weiteren sind die Wahrscheinlichkeitsfunktionen $P_A(a)$ und $P_N(r)$ enthalten, um die Verteilung der einzelnen Empfänger a, r von Alice und von den der anderen Sender in den Observationen zu modellieren. Da das Sammeln von Nachrichten aufwendig ist, nehmen wir an, dass in jeder Runde nur ein kleiner Teil (d.h. $b \ll u$) aller möglichen Empfänger kontaktiert werden.

- Wir ermitteln den Erwartungswert der theoretischen Anonymitätsschranke über die erwartete minimale Anzahl von Observationen, bis der Mix die Anonymität von Alice nicht mehr gewährleistet.[4] Diese ähnelt Shannons Unicity-Distance [Sha49].
- Zusätzlich analysieren wir den erwarteten Rechenaufwand, den diese Deanonymiserung erfordern würde. Dadurch wird ersichtlich, welche Angreifer in der Praxis diese Ressourcen aufbringen können.

[3] Im Mixmaster wird ein Paket in einer Runde zufällig weitergeleitet, oder zurückgehalten. Die Weiterleitung von Paketen ist daher indeterministisch.

[4] Diese Schranke quantifziert wie die Unicity-Distance die Information, aber nicht den Rechenaufwand für eine Deanonymisierung einer Kommunikationsbeziehung.

Diese theoretische und praktische Analyse basiert auf dem Hitting-Set Angriff (HS-Attack) [KP04]. Der Angriff erhält, wie in Abbildung 2 illustriert, die Observationen als Eingabe und leitet daraus die möglichen *Hypothesen* für die Menge der m Freunde von Alice ab. Alice ist deanonymisiert, wenn der HS-Attack genau eine Hypothese liefert. Es wurde in [KAPR06] bewiesen, dass der HS-Attack die minimale Anzahl an Observationen für eine Deanonymisierung benötigt. Die erwartete minimale Anzahl an Beobachtungen, um Alice in einer gegebenen Mix Konfiguration eindeutig zu identifizieren, entspricht dem Erwartungswert der Anonymitätsschranke. Allerdings erfordert der HS-Attack das Lösen eines NP-vollständigen Problems.

1.1 Beitrag

Der Beitrag dieser Dissertation kann in drei Bereiche unterteilt werden. Erstens werden wir den Erwartungswert der theoretischen Schranke der Anonymität für ein beliebiges Subjekt, das wir Alice nennen, durch eine geschlossene Formel approximieren. Diese Schranke ist von der Konfiguration des Chaum Mixes abhängig. Unsere Aprroximation zeigt, dass die Konfiguration des Mixes für gewöhnlich nur einen polynomiellen Einfluss auf die theoretische Schranke der Anonymität hat.

Zweitens bestimmen wir anhand eines konkreten Angriffs den mittleren Rechenaufwand, um die Freunde von Alice zu identifizieren. Dies zeigt, dass die Identifizierung für viele realistische Parameter im Mittel effizient berechenbar ist, obwohl es das Lösen eines NP-vollständigen Problems erfordert.

Drittens schlagen wir eine Erweiterung unserer Analysen auf indeterministische Mix Strategien vor, in der fehlerhafte Observationen möglich sind. Es wird gezeigt, dass Alice mit hoher Wahrscheinlichkeit deanonymisiert werden kann, wenn die Fehlerwahrscheinlichkeit der Observationen unter einer bestimmten Schranke liegt. Strategien, die durch die Induzierung fehlerhafter Observationen die Anonymität stärken wollen, müssen daher diese Schranke berücksichtigen.

2 Theoretische Schranke der Anonymität

Die theoretische Schranke der Anonymität bestimmt, ab wann der Chaum Mix die Anonymität eines Senders nicht mehr schützen kann. Wir messen diese Schranke über die minimale Anzahl an Observationen, die ein Angriff benötigt, um die Freunde von Alice eindeutig zu identifizieren. Die Schranke ist theoretisch, da sie die für einen solchen Angriff erforderliche Zeit- und Speicherkapazitäten nicht berücksichtigt.

Sei ${}_A\mathcal{H} = \{a_1, \ldots, a_m\}$ die Menge aller Freunde, die Alice während der Observationen des Angreifers wiederholt kontaktiert und $m = |{}_A\mathcal{H}|$ die Anzahl dieser Freunde. In der Basis Analyse betrachten wir den Fall, dass Alice während des Angriffs nur Empfänger in ${}_A\mathcal{H}$ kontaktiert. Diese Betrachtung wird in Kapitel 4 um fehlerhafte Observationen erweitert.

2.1 Hitting-Set Angriff

Wie in Abbildung 2 illustriert, bestimmt ein Angriff aus Observationen Hypothesen für die Menge der Freunde von Alice. Da jede Observation mindestens einen Freund von Alice enthält, ist jede Hypothese ein Hitting-Set der Größe m. Ein *Hitting-Set* ist in unserem Kontext eine Menge, die sich mit jeder Observation des Angreifers schneidet. Der Hitting-Set ist ein *Minimal-Hitting-Set*, wenn keine echte Teilmenge davon ein Hitting-Set ist. Ein Hitting-Set wird zum *Unique Minimum-Hitting-Set* (UMHS), wenn alle anderen Hitting-Sets eine höhere Kardinalität haben.

Der Hitting-Set Angriff (HS-Attack) [KP04] sammelt wiederholt Observationen und berechnet die Hitting-Sets auf alle gesammelten Observationen, bis er ein UMHS findet. Dieser UMHS identifiziert eindeutig die Freunde von Alice. Dieser Angriff nutzt den Umstand aus, dass die Sender und Empfänger in einer offenen Nutzergruppe sich fortlaufend ändern, während die Observationen immer einen Freund von Alice enthalten. Bei hinreichend vielen Observationen wird die Menge ${}_A\mathcal{H}$ aller Freunde von Alice daher zum UMHS in diesen Observationen[5]. Es wurde in [KAPR06] bewiesen, dass der HS-Attack die minimale Anzahl an Observationen benötigt, um die Freunde von Alice eindeutig zu identifizieren. Das Bestimmen des Unique Minimum-Hitting-Set ist ein NP-vollständiges Problem [GJ90].

2.2 Approximation der Anonymitätsschranke

Der Erwartungswert für die minimale Anzahl an Observationen um Alice zu deanonymisieren ist, wie in Abbildung 2 illustriert, abhängig von der Konfiguration des Mixes. Neben den Parametern u, b, m enthält diese die Wahrscheinlichkeitsfunktionen $P_A(a)$ und $P_N(r)$. Die Funktion $P_A(a)$ modelliert für jeden Freund a die Wahrscheinlichkeit, dass dieser von Alice kontaktiert wird. Für jeden möglichen Empfänger r beschreibt $P_N(r)$ hingegen die Wahrscheinlichkeit, dass r in einer Observation vorkommt, weil es von eines der anderen $(b-1)$ Sendern kontaktiert wird. Sei p der minimaler Wert von $P_A(a)$, für alle $a \in {}_A\mathcal{H}$ und P_N, der maximale Wert von $P_N(r)$ für alle möglichen Empfänger r. Sei ferner der Zusammenhang $P_N = 1 - (\frac{u-1}{u})^{b-1}$ gegeben[6].

Die Arbeit stellt folgende Approximation für den Erwartungswert der theoretischen Anonymitätsschranke im Bezug auf die Konfiguration des Mixes auf:

$$E(T_{2\times e}) \approx \left(\frac{1}{p}(\ln m + \gamma) + \frac{1}{p}\ln\ln m\right)\left(\frac{u-(m-1)}{u}\right)^{1-b} \quad , \tag{1}$$

wobei $\gamma \approx 0,57721$ die Euler-Mascheroni Konstante ist. Dabei ist u für gewöhnlich wesentlich größer als m, so dass der letzte Faktor in (1) durch $\frac{1}{1-(b-1)(\frac{m-1}{u})}$ approximiert werden kann, so dass $E(T_{2\times e})$ polynomiell ist. Daraus wird ersichtlich, dass der Chaum

[5] Es gibt theoretische Ausnahmefälle, in der dieses nicht erfüllt ist. Z.B. wenn alle Sender immer den gleichen Empfänger kontaktieren.

[6] Die Herleitung kann der Dissertation entnommen werden.

Mix die Anonymität eines Senders nur für eine polynomielle Anzahl an Kommunikationen schützt. Änderungen an der Konfiguration des Mixes würden nur eine polynominelle Wirkung auf den Erwartungswert der theoretischen Anonymitätsschranke haben.

Evaluation Abbildung 3 veranschaulicht die Anonymitätsschranke für verschiedene Konfigurationen des Mixes. Es vergleicht die Approximation ($E(T_{2\times e})$) mit der Erwarteten Anzahl an simulierten Observationen, die der HS-Attack für die Deanonymisierung von Alice benötigt (HS). Die Observationen werden wie in Abbildung 2 zufällige generiert. Dabei haben wir vereinfachend eine gleichförmige Wahrscheinlichkeitsfunktion $P_N(r) = 1 - (\frac{u-1}{u})^{b-1}$ gewählt[7]. Alice kontaktiert ihre Freunde nach der Zipf(m, α) Verteilung $P_A(i) = \frac{i^{-\alpha}}{\sum_{l=1}^{m} l^{-\alpha}}$, für $a_i \in {}_A\mathcal{H}$, $i = \{1, \ldots, m\}$.[8] Die Zipf-Verteilung ähnelt für $\alpha \approx 1$ der Verteilung des E-Mail Verkehrs eines Nutzers [BCF+99, AH02]. Die Anonymitätschranken in Abbildung 3 bestätigen, dass der Chaum Mix die Anonymität nur für eine relativ kleine Anzahl von Runden gewährleistet. Im rechten unteren Bild steigt die Anzahl der Beobachtungen exponentiell mit α. Das ist kein Widerspruch zum Kapitel 2.2, da der minimaler Wert p der Funktion $P_A(i)$ mit wachsendem α exponentiell abnimmt.

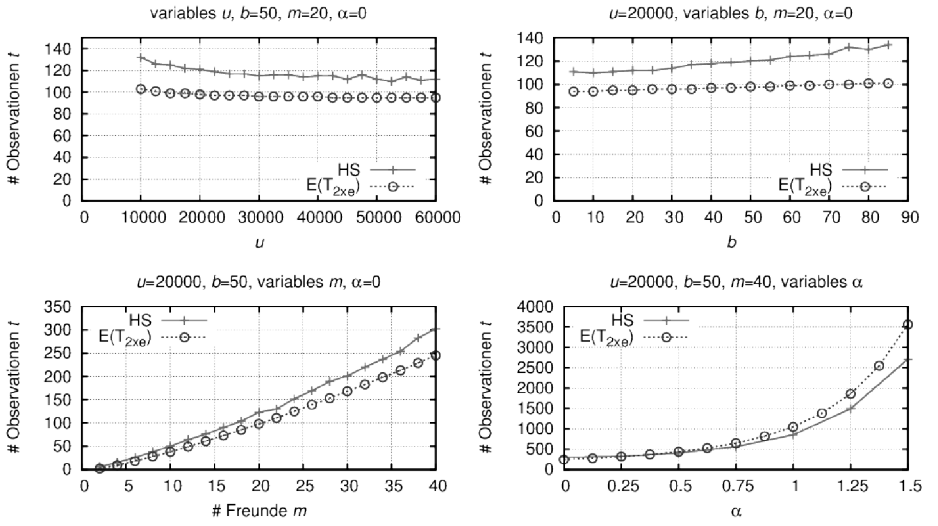


Abbildung 3: Empirischer und theoretischer Erwartungswert der Anonymitätsschranke: Empirisch durch Simulation mit HS-attack (HS) und theoretisch durch Approximation ($E(T_{2\times e})$).

[7] Es können auch komplexere Verteilungen für $P_N(r)$ betrachtet werden. Allerdings wäre die mathematische Analyse unnötig umständlicher.

[8] Da Alice durch ihr Verhalten einen direkten Einfluss auf ihre Anonymität hat, wird die Wirkung ihrer Kommunikation detailliert analysiert.

3 Rechenaufwand der Deanonymisierung

Der HS-Attack und die damit ermittelte Anonymitätsschranke wurde als eine rein theoretische Analyse betrachtet. Um ein UMHS zu identifizieren, wird ein Algorithmus verwendet, der alle möglichen Hitting-Sets der Größe m in den gesammelten Observationen berechnet. Wenn genügend viele Observationen vorhanden sind, liefert dieser Algorithmus genau ein Hitting-Set, der zugleich ein UMHS ist, vergl. [KP04]. Da es anfänglich[9] $\binom{u}{m}$ Hitting-Sets gibt, ist die Komplexität des Algorithmus $O(\binom{u}{m})$, was für realistische Gruppengrößen inpraktikabel ist. Diese Dissertation stellt einen neuen Algorithmus vor, der für viele realistische Konfigurationen des Mixes das UMHS effizient identifizieren kann.

3.1 ExactHS Algorithmus

In dieser Arbeit haben wir den *ExactHS* Algorithmus entworfen, der für eine gegebene Menge an Observationen nur die Minimal-Hitting-Sets bis zu einer gegebenen Kardinalität (z.B. m) berechnet. Da jedes Hitting-Set, das nicht minimal ist, eine Obermenge eines Minimal-Hitting-Sets ist, führt der ursprünglicher Algorithmus des HS-Attack durch die Betrachtung aller Hitting-Sets der größe m zu einem erheblich höheren Rechenaufwand. Wir beweisen in dieser Dissertation, dass es maximal b^m Minimal-Hitting-Sets gibt und die Worst-Case Laufzeit des ExactHS proportional dazu ist, während dessen Speicherkomplexität nur linear ist. Zum Beispiel ist bei einer Mixkonfiguration mit $u = 400$ ($u = 10000$) Empfängern, einer Anonymitätsmenge der Größe $b = 10$ ($b = 50$) bei einer Anzahl von $m = 10$ ($m = 20$) Freunden, die Laufzeitkomplexität des ursprünglichen Algorithmus 10^{19} (10^{61}), während sie beim ExactHS 10^{10} (10^{34}) ist. Die Verwendung des ExactHS, um ein UMHS zu identifizieren liefert bezüglich der minimalen Anzahl an erforderlichen Observationen dasselbe Ergebnis, wie der ursprüngliche Algorithmus des HS-Attack. Der HS-Attack unter der Verwendung des ExactHS zur Bestimmung der Minimum-Hitting-Sets stellt einen effizienteren Entwurf des HS-Attack dar. Daher beziehen wir uns in dem restlichen Text nur noch auf die effizientere Version, wenn wir den Begriff HS-Attack verwenden.

3.2 Mittlere Komplexität des ExactHS

Der ExactHS ist insbesondere praktisch verwendbar, weil seine mittlere Laufzeitkomplexität in vielen Fällen signifikant geringer als seine Worst-Case Komplexität ist. Die Dissertation beweist folgende mathematische Zusammenhänge für die mittlere Laufzeitkomplexität des ExactHS:

- Wenn die Wahrscheinlichkeit, dass ein Empfänger von irgendeinem Sender in einer Runde kontaktiert wird, der nicht Alice ist, höchstens $\frac{1}{m^2}$ ist (d.h. $P_N \leq \frac{1}{m^2}$), dann ist die mittlere Laufzeitkomplexität linear.

[9] Wenn noch keine Observation gesammelt wurde.

- Wenn $P_N \leq \frac{1}{m}$ ist, und Alice ihre Freunde ähnlich einer Zipf(m, α) Verteilung kontaktiert (d.h. $P_A(i) = \frac{i^{-\alpha}}{\sum_{l=1}^{m} l^{-\alpha}}$, für $a_i \in {}_A\mathcal{H}$, $i = \{1, \ldots, m\}$), dann konvergiert die mittlere Laufzeitkomplexität mit steigendem α gegen ein Polynom. Tendenziell führt eine ungleichförmige Verteilung der Nachrichten von Alice an ihre Freunde zu einer Verringerung der mittleren Laufzeitkomplexität.

Es wurde in der Literatur empirisch bestätigt, dass der E-Mail Verkehr und der Datenverkehr im Internet durch eine Zipf-Verteilung modelliert werden kann [BCF+99, AH02]. Daher ist der verbesserte HS-Attack insbesondere bei vielen realistischen Verteilungen effizient durchführbar. Zum Beispiel liegt die Worst-Case Komplexität von ExactHS für $b = 50, m = 40$ bei $O(50^{40}) \approx O(10^{68})$. Die in dieser Dissertation mathematisch und simulativ[10] ermittelte Komplexität führt hingegen zu den signifikant geringeren Werten in Tabelle 1. Es ist auch erkennbar, dass mit zunehmendem α die Komplexität weiter sinkt.

α	Theoretische Anzahl	Empirische Anzahl
0.0	1.8×10^7	1.7×10^6
0.5	1.0×10^5	1.5×10^5
1.0	8.7×10^3	2×10^4
1.5	2.2×10^3	1.7×10^3

Tabelle 1: Evaluierter Mengen zur Identifizierung des UMHS durch ExactHS: Für Mixparameter $u = 20000, b = 50, m = 40$ und Zipf(m, α) Verteilung der Nachrichten von Alice und $P_N < \frac{1}{m}$.

4 Erweiterte Analyse mit fehlerhaften Observationen

Bei der Verwendung des HS-Attack wird angenommen, dass die Observationen des Angreifers fehlerfrei sind. Die Dissertation zeigt, dass der HS-Attack adaptiert werden kann, um die Freunde von Alice mit einer hohen Wahrscheinlichkeit zu identifizieren, wenn die Fehlerrate der Observationen eine bestimmte Schranke unterschreitet. Eine Observation ist *fehlerhaft*, wenn sie entgegen der Erwartung des Angreifers keinen Freund von Alice enthält. Fehlerhafte Observationen können durch das Senden von Scheinnachrichten, oder unvollständige Beobachtungen der Anonymitätsmengen entstehen. Zum anderen können diese bei Mix Varianten auftreten, die Nachrichten indeterministisch weiterleiten, wie z.B. Pool-Mixe [DDM03]. Bei Pool-Mixen (wie Mixmaster [DDM03]) besteht eine Wahrscheinlichkeit p_{er} dass die Weiterleitung einer Nachricht auf eine spätere Runde aufgeschoben wird. Ein Angreifer, der die Empfängermenge nur in der Runde des Pool-Mixes beobachtet, in der Alice eine Nachricht sendet, wird daher mit der Wahrscheinlichkeit p_{er} eine fehlerhafte Observation erhalten.

[10] Dafür wurden entsprechend der Konfiguration des Mixes zufällige Observationen erzeugt und der HS-Attack auf diese angewendet, bis alle Alices Freunde identifiziert wurden. Diese Simulation wurde mehrmals wiederholt, bis 95% der Ergebnisse sich um maximal 5% vom empirischen Mittelwert unterschieden.

Die Dissertation zeigt den folgenden Zusammenhang: Wenn die Wahrscheinlichkeit einer fehlerhaften Observation die Ungleichung

$$p_{er} < \frac{1}{(\frac{1}{p}\ln\ln m)\left(\frac{u-(m-1)}{u}\right)^{1-b}+1} \tag{2}$$

erfüllt, dann konvergiert die Wahrscheinlichkeit, dass Alices Freunde durch den HS-Attack identifiziert werden kann, mit steigender Anzahl an Observationen gegen 1. In (2) is p der minimale Wert der Wahrscheinlichkeitsfunktion $P_A(a)$, für $a \in {}_A\mathcal{H}$.

Die Verwendung von Pool-Mixen schützt somit nicht gegen den erweiterten HS-Attack, wenn die Fehlerwahrscheinlichkeit von Observationen (2) erfüllt. Höhere Werte für p_{er} führen jedoch zu stärkeren Verzögerungen der Pakete. Allerdings könnte, selbst wenn (2) nicht erfüllt wird, die möglichen Hypothesen für die Freunde von Alice durch den HS-Attack stark einschränkt werden.

5 Zusammenfassung

Diese Arbeit [Pha13] beweist, dass der Chaum Mix einen Sender im Mittel nur für eine polynomielle Anzahl an Kommunikationen schützen kann. Durch eine Verbesserung der Effizienz des ursprünglichen HS-Attack und der Analyse der mittleren Laufzeit des Angriffs können wir zeigen, dass die Deanonymisierung eines Senders (d.h. die Identifizierung seiner Empfänger) in vielen realistischen Fällen sogar effizient möglich ist. Aus diesen Ergebnissen folgern wir, dass der Chaum Mix (der deterministisch Nachrichten weiterleitet), keinen starken Schutz gegenüber dem HS-Attack bietet.

Wir haben den HS-Attack daher erweitert, um auch Mix Varianten betrachten zu können, die indeterministisch Nachrichten weiterleiten, wie dem Pool-Mix. Dieser Indeterminismus führt zu fehlerhaften Observationen. Wir haben gezeigt, dass der HS-Attack die Freunde von Alice mit einer hohen Wahrscheinlichkeit identifizieren kann, wenn die Fehlerrate der Observationen einen bestimmten Grenzwert unterschreitet. Wenn dieser Grenzwert berücksichtig wird, können indeterministische Mix Varianten einen besseren Schutz vor dem HS-Attack bieten, als der Chaum Mix. Allerdings führt der Indeterminismus zu zusätzlichen Verzögerungen des Datenverkehrs und einer Aufweichung des Angreifermodells, im Vergleich zum Chaum Mix. Die Verzögerung erschwert insbesondere die Erkennung von aktiven Angriffen, wie z.B. das Blocken von $(b-1)$ Nachrichten vor dem Mix[11] [Cha81, SDS03], da Verzögerungen konzeptbedingt auch im Normalfall auftreten. Im Gegensatz dazu ermöglicht die deterministische Arbeitsweise des Chaum Mixes die Erkennung aktiver Angriffe durch die CUVE [KP06] Anforderungen.

Die vorliegende Arbeit zeigt die Schwächen bestehender Mix Varianten auf und motiviert daher die Forschung, nach neuen Techniken zu suchen, die eine Balance zwischen den genannten Vorteilen und Nachteilen herstellen.

[11] Wenn nur noch eine Nachricht den Mix passiert, dann ist dessen Sender und Empfänger nicht mehr anonym.

Literatur

[AH02] Lada A. Adamic und Bernardo A. Huberman. Zipf's Law and the Internet. *Glottometrics*, 3:143 – 150, 2002.

[BCF+99] Lee Breslau, Pei Cao, Li Fan, Graham Phillips und Scott Shenker. Web Caching and Zipf-like Distributions: Evidence and Implications. In *Proceedings of Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM '99*, Jgg. 1, Seiten 126 – 134. IEEE, 1999.

[Cha81] David L. Chaum. Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. *Communications of the ACM*, 24(2):84 – 88, 1981.

[DD08] George Danezis und Claudia Diaz. A Survey of Anonymous Communication Channels. Bericht MSR-TR-2008-35, Microsoft Research, 2008.

[DDM03] George Danezis, Roger Dingledine und Nick Mathewson. Mixminion: Design of a Type III Anonymous Remailer Protocol. In *Proceedings of the 2003 IEEE Symposium on Security and Privacy*, Seiten 2 – 15. IEEE, 2003.

[EY09] Matthew Edman und Bülent Yener. On Anonymity in an Electronic Society: A Survey of Anonymous Communication Systems. *ACM Computing Surveys*, 42(1):1 – 35, 2009.

[GJ90] Michael R. Garey und David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, 1990.

[KAPR06] Dogan Kesdogan, Dakshi Agrawal, Vinh Pham und Dieter Rauterbach. Fundamental Limits on the Anonymity Provided by the Mix Technique. In *Proceedings of the 2006 IEEE Symposium on Security and Privacy*, Seiten 86 – 99. IEEE, 2006.

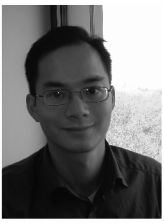
[KP04] Dogan Kesdogan und Lexi Pimenidis. The Hitting Set Attack on Anonymity Protocols. In *Information Hiding*, Jgg. 3200 of *LNCS*, Seiten 326 – 339. Springer, 2004.

[KP06] Dogan Kesdogan und Charles Palmer. Technical Challenges of Network Anonymity. *Computer Communications*, 29(3):306 – 324, 2006.

[Pha13] Dang Vinh Pham. *Towards Practical and Fundamental Limits of Anonymity Protection*. Dissertation, University of Regensburg, November 2013.

[SDS03] Andrei Serjantov, Roger Dingledine und Paul F. Syverson. From a Trickle to a Flood: Active Attacks on Several Mix Types. In *Information Hiding*, Jgg. 2578 of *LNCS*, Seiten 36 – 52. Springer, 2003.

[Sha49] Claude Shannon. Communication Theory of Secrecy Systems. *Bell System Technical Journal*, 28:656 – 715, 1949.



Vinh Pham wurde 1981 in Saigon (Vietnam) geboren. Er erhielt 2006 das Diplom für Informatik an der Rheinisch-Westfälischen Technischen Hochschule Aachen (RWTH-Aachen). Ergebnisse der Diplomarbeit führten zu der Publikation "Fundamental limits on the anonymity provided by the MIX technique" bei der IEEE Security&Privacy 2006. Nach seinem Abschluss war er wissenschaftlicher Mitarbeiter im RWTH UMIC Exzellenz Cluster. Das Forschungsinteresse für den Schwerpunkt Security and Privacy führte ihn 2008 an die Universität Siegen und dann Regensburg. Seine Promotion schloss er 2013 in Regensburg ab, wo er bis heute seine Forschungsrichtung in der Lehre und in Projekten weiterverfolgt.