fhwedel
UNIVERSITY OF APPLIED SCIENCES

Seminar IT-Sicherheit

Traffic-Analyse und Deanonymisierung

Eingereicht am:

17. November 2015

Jens Begemann
Matr.-Nr.: Inf 101419
Pestalozzistrße 51
25421 Pinneberg
Phone: (0176) 237 346 21
E-Mail: inf101419@fh-wedel.de

Betreut von:
Prof. Dr. Gerd Beuster
Feldstraße 143
22880 Wedel
Phone: (04103) 80 48 - 38
E-Mail: gb@fh-wedel.de

"Das Recht auf Anonymität ist an sich so selbstverständlich, dass man darüber nicht schreiben oder sprechen müsste."[1]

[1]Die Landesbeauftragte für den Datenschutz Niedersachsen (2015)

Inhaltsverzeichnis

1 Einleitung **1**

2 Identität und Anonymität **3**
2.1 Anonymisierungsnetzwerk Tor 4

3 Traffic-Analyse **6**
3.1 Deanonymisierung von Tor Nutzern 6
3.1.1 NetFlow allgemein 7
3.1.2 Voraussetzungen für die NetFlow-Analyse 8
3.1.3 Durchführung des Angriffs 9
3.1.4 Korrelation der NetFlow Daten 11

4 Zusammenfassung **15**

Abbildungsverzeichnis **16**

Tabellenverzeichnis **17**

Literaturverzeichnis **18**

1 Einleitung

Jede Aktion in der digitalen Welt hinterlässt heutzutage vielfältige Spuren - genannt Metadaten. Der Besuch einer Website, das Schauen eines Videos auf einem Videoportal, das Kommentieren eines Beitrags in einem sozialen Netzwerk oder einfach nur der Austausch von Nachrichten zwischen einer oder mehreren Personen zusammen mit deren Standort und der Uhrzeit. All diese Aktionen haben gemeinsam, dass sie in der Regel für Dritte nachvollziehbar und auswertbar sind und mit der Wiedereinführung der Vorratsdatenspeicherung langfristig gespeichert werden [1].

Während für viele Menschen die Verbreitung und Speicherung dieser Informationen unangenehm aber vermeintlich unbedenklich sind, gibt es auch Gruppen die sich davor schützen müssen. Dazu zählen in erster Linie Whistleblower, Aktivisten, Oppositionelle und häufig auch Journalisten. Der Nachweis einer Kommunikation zwischen Mitgliedern aus diesen Gruppen untereinander kann unter Umständen ernsthafte Folgen haben, wie Inhaftierung oder im schlimmsten Fall der Tod [2].

In der Bundesrepublik Deutschland hat jeder Mensch ein Recht auf Privatsphäre. Die Erhebung, Verarbeitung und Nutzung personenbezogener Daten ist nur zulässig sofern sie durch eine Rechtsvorschrift angeordnet wurde oder der Betroffene damit einverstanden ist [3]. Ausgehend von der gültigen Rechtsprechung kann der Wunsch nach Anonymität demnach kein Verbrechen sein, sondern wird jedem als Grundrecht zugestanden. Vielmehr wird im Telemediengesetz jedem Internetnutzer das Recht eingeräumt sich anonym durch das Netz zu bewegen [4].

Trotz der eindeutigen Rechtslage werden von vielen Seiten personenbezogene Daten erhoben und verarbeitet, in erster Linie von sozialen Netzwerken und kostenlosen Webangeboten. Neben den vermeintlich harmlosen Webseitenbetreibern die mit diesen Daten zielgenau Werbung an den Nutzer bringen möchten und detaillierte Statistiken über das Nutzerverhalten erstellen, um ihr Angebot besser auszurichten zu können, wer-

den diese Daten allerdings auch von Behörden und mitunter kriminellen Organisationen erfasst.

Viele Nutzer des Internets haben ein sehr starkes Interesse sich anonym zu bewegen und miteinander unerkannt zu kommunizieren. Journalisten und Whistleblower oder politisch Verfolge sind an dieser Stelle die prominentesten Vertreter. Um anonym miteinander kommunizieren zu können müssen zwangsweise Anonymisierungsdienste wie das Tor-Netzwerk eingesetzt werden. Jedoch können diese Dienste auch von Kriminellen eingesetzt werden. Die Behörden haben somit auch ein berechtigtes Interesse im Rahmen der Strafverfolgung bestimmte Nutzer zu identifizieren.

Während in dem Zusammenhang mit Traffic-Analyse meist von der Auswertung von Webseiten- und Dienst-Benutzungen die Rede ist, werden in diesem Seminar Methoden betrachtet, die zur Identifizierung von Benutzern eines Anonymisierungsdienstes wie dem Tor-Netzwerk eingesetzt werden.

Im Folgenden wird ein praktisches Verfahren beschrieben mit dem versucht wird, die Nutzer des Anonymisierungsdienstes Tor zu deanonymisieren.

2 Identität und Anonymität

Anonymität bedeutet, dass eine Person oder eine Gruppe nicht identifiziert werden kann [5]. Eine Identifizierung, also die Festellung einer bestimmten Identität, kann durch das Erlangen von Informationen über eine Person erfolgen. Zu diesen Informationen zählen unter anderem sein gesetzlicher Name oder Pseudonyme, Adressen (postalisch und IP-Adressen), Verhaltensmuster und viele mehr [6].

Die einwöchige Auswertung von Metadaten nur eines einzelnen Mobiltelefons genügt, um ein detailierstes Persönlichkeitsprofil der betroffenen Person zu erstellen. Dies beinhaltet nicht nur Bewegungsprofile sondern offenbart nahezu das gesamte soziale Umfeld des Betroffenen [7]. Um sich, zumindest teilweise, dieser Erfassung zu entziehen können verschiedenen Anonymisierungsdienste eingesetzt werden.

Bei der Betrachtung von Anonymisierungsnetzwerken wird zwischen low- und high-latency Netzen (Mixminion[1], Mixmaster[2]) unterschieden. Während low-latency Netze wie Tor das Surfen im Internet, Instant-Messaging und je nach verfügbarer Bandbreite auch das Streamen von Videos ermöglichen, ist dies bei high-latency Netzen nicht möglich. Die Datenpakete werden in diesen Netzwerken für die Dauer von mehreren Stunden bis hin zu Tagen verzögert. Dies macht eine Korrelation von ein- und ausgehenden Datenpaketen nahezu unmöglich. Bei der Verwendung von Anonymisierungsdiensten wie dem Tor Netzwerk wird die Traffic-Analyse mit dem Ziel eingesetzt, die Teilnehmer gegenüber einer Ressource zu deanonymisieren.

[1] http://www.mixminion.net/
[2] http://mixmaster.sourceforge.net/

2.1 Anonymisierungsnetzwerk Tor

Tor ist ein Anonymisierungsnetzwerk welches sich aus vielen Servern, die meist von Freiwilligen betrieben werden zusammensetzt. Ziel des Netzwerkes ist es, die Privatsphäre der Nutzer zu schützen und die Kommunikation gegenüber Dritten zu verschleiern. Weiterhin kann es dafür genutzt werden, um regionale Zensurmaßnahmen zu umgehen. Tor kann seine Nutzer vor tracking schützen und bietet die Möglichkeit anonyme Dienste bereitzustellen, die nur innerhalb des Netzwerkes erreichbar sind [8].

Ein Nutzer verbindet sich mit dem Tor-Netzwerk indem zuerst eine Liste mit verfügbaren Relays aus einem öffentlich bekannten Directory-Service angefragt wird. Anschließend wird eine verschlüsselte TLS-Verbindung mit mindestens drei Relays aufgebaut. Dadurch wird gewährleistet, dass ein Relay zu keinem Zeitpunkt gleichzeitig Sender und Empfänger kennt. Während sowohl die Verbindungen zu dem Directory-Service als auch innerhalb des Tor Netzwerkes über die TLS-Transportverschlüsselung gesichert sind, ist der ausgehende Traffic unverschlüsselt (Abbildung 2.1).

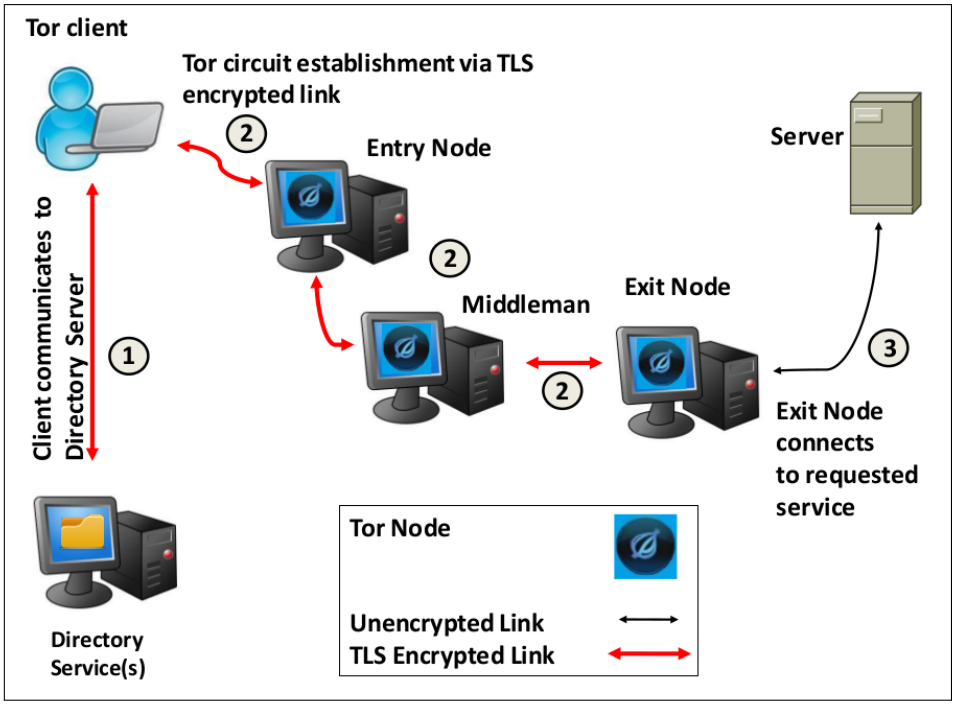


Abbildung 2.1: Tor Netzwerk

Die Verwendung des Tor-Netzwerkes bringt allerdings auch Risiken mit sich. Der Betreiber einer Tor-Exit-Node kann den gesamten Traffic, der darüber abgewickelt

wird, mitlesen ohne das dafür ein richterlicher Beschluss notwendig ist. Das bedeutet in dem Fall von unverschlüsselten Daten den kompletten Zugriff auf deren Inhalt. Unverschlüsselte Anmeldevorgänge mit Benutzername und Passwort sind nur ein Beispiel. Bedingt durch das Freiwilligenprinzip des Tor Netzwerkes können Relays und speziell Exit-Nodes auch unbemerkt von Kriminellen und/oder Behörden betrieben werden. Eine zusätzliche Inhaltsverschlüsselung der Daten ist somit unerlässlich, denn ohne sind die Gefahren durch die man sich mittels Tor zu schützen versucht größer als ohne die Verwendung des Dienstes [9].

3 Traffic-Analyse

Der Begriff Traffic-Analyse bezieht sich in diesem Seminar auf die Auswertung von IP-Datenströmen, die beim Austausch von Nachrichten bzw. Daten die zwischen zwei Kommunikationsteilnehmern anfallen. Um eine Traffic-Analyse durchzuführen, müssen die Nachrichten üblicherweise auf der Strecke zwischen den Teilnehmern abgefangen und analysiert werden. In diesem Fall ist die Auswertung von *Flow* Informationen, die beim Passieren der Daten durch einen Router oder Layer-3-Switch anfallen, ausreichend. Der Ablauf einer Kommunikation besteht aus mehreren Datenpaketen. Die Analyse dieses Datenstroms hat das Ziel durch statistische Korrelation bestimmte Muster in der Kommunikation zu finden, die zuvor durch Bandbreitenmanipulation injiziert wurden (*fingerprinting*) um bestimmte Teilnehmer durch ihre IP-Adressen zu identifizieren. Der Inhalt der Nachrichten kann verschlüsselt sein, was keinen Einfluss auf die in diesem Seminar behandelten Methoden hat.

Im Jahr 2007 wurde durch Murdoch et al. [10] ein Modell zur Traffic-Analyse und Deanonymisierung von Tor Nutzern durch Netflow Daten beschrieben. In diesem Seminar wird ein praktischen Angriff nach Chakravarty et al. [11] aus dem Jahr 2013 beschrieben, der auf dem Modell von Murdoch basiert.

3.1 Deanonymisierung von Tor Nutzern

Die Architektur des Tor Netzwerkes erlaubt eine Identifizierung der Nutzer sofern ein Angreifer in der Lage ist alle Ein- und Austrittspunkte des Netzwerks zu überwachen. Die Sicherheit des Tor Netzwerkes besteht demnach in der Vielzahl von Servern. Je mehr Server von Freiwilligen auf der ganzen Welt betrieben werden, umso schwieriger wird es für Angreifer jeden dieser Knoten zu überwachen. Durch die Möglichkeit der Netzbetreiber und Behörden großflächig die Kommunikation zu überwachen, sollte eine Route demnach durch möglichst verschiedene Länder bzw. Hoheitsbereiche erfolgen. Der

Aufwand eine flächendeckende Überwachung gegen ein solches Netzwerk zu etablieren erfordert enorme Ressourcen und lässt sich nur schwer realisieren.

Im Folgenden wird ein praktischer Angriff beschrieben Nutzer des Tor Netzwerkes durch die Auswertung von Informationen aus dem Netzwerkmonitoring und der Bandbreitenüberwachung (*Flows*) zu deanonymisieren. Es wird davon ausgegangen, dass der Angreifer in der Lage ist, wenige aber große Knotenpunkte des Internets zu überwachen. Weiterhin ist er in der Lage von Ziel der anonymen Verbindung ausgehend den Datenstrom durch Bandbreitenmanipulation zu verändern.

3.1.1 NetFlow allgemein

Netflow ist ein passives Messverfahren bei dem der Datenverkehr beobachtet werden kann ohne ihn zu beeinflussen. Die NetFlow-Daten werden von Routern und Layer-3-Switches erstellt und lassen sich zentral auswerten, somit ist für den Angriff keine Installation von Netzwerkequipment notwendig. Die meisten Router und Switches in Rechenzentren stellen diese Flows[1] bereit. Über UDP können die Daten von einem Netflow-Kollektor gesammelt und ausgewertet werden. Die für die Korrelation wichtigsten Eigenschaften eines Flows (Abbildung 3.1) sind neben dem Start- und Endzeitpunkt der Aufzeichnung die Quell- und Zieladresse sowie der verwendete Port und die Anzahl Pakete (bzw. Bytes), die in dem Zeitraum übertragen wurden.

Start	End	Sif	SrcIPaddress	SrcP	DIf	DstIPaddress	DstP	P	Fl	Pkts	Octets
0606.23:59:06.616	0606.23:59:36.660	65535	192.168.0.20	50000	2	213.163.65.50	54089	6	0	3	156
0606.23:59:06.616	0606.23:59:36.660	2	213.163.65.50	54089	65535	192.168.0.20	50000	6	0	3	1914
0606.23:59:29.420	0606.23:59:30.572	2	71.58.107.145	42259	65535	192.168.0.20	50000	6	2	10	3961
0606.23:59:29.420	0606.23:59:30.612	65535	192.168.0.20	50000	2	71.58.107.145	42259	6	2	9	3578
0606.23:59:33.396	0606.23:59:33.396	65535	127.0.0.1	55171	1	127.0.0.1	10002	17	0	1	1492

Abbildung 3.1: NetFlow Datensatz

Die Flows werden zur Laufzeit im Speicher der Geräte angelegt und mit jedem Paket, welches den Router passiert, aktualisiert. Jeder Flow ist mit zwei Timern verknüpft. Ein *active* Timer, welcher den Flow als aktiv markiert wenn in seinem Zeitraum Daten übertragen werden. Der zweite Timer legt die *inactive* Dauer fest nachdem ein Flow als inaktiv markiert wird wenn in diesem Zeitraum keine Daten übertragen wurden. Nach Ablauf der Timer wird der Flow Record vom Router auf ein persistenten Speicher übertragen. Von dort aus können die Flows ausgewertet werden.

[1] Cisco NetFlow [12], Huawei NetStream [13], Alcatel Lucent sFlow [14]

3.1.2 Voraussetzungen für die NetFlow-Analyse

Damit durch die Informationen aus den Flows eine Identifizierung von anonymen Verbindungen erfolgen kann, muss der Angreifer in der Lage sein die Flow-Daten an den Ein- und Austrittspunkten der Kommunikation abzugreifen. Dazu müssen die Flows vom Server zum Austrittspunkt und vom Eintrittspunkt zu den Clients korreliert werden (Abbildung 3.2). Im Gegensatz zu vielen bisherigen Angriffsmodellen gegen das Tor-Netzwerk, muss der Angreifer keinen direkten Zugriff auf die Ein- und Ausgangsknoten haben, sondern kann die Flows an zentraler Stelle eines Providers abfangen.

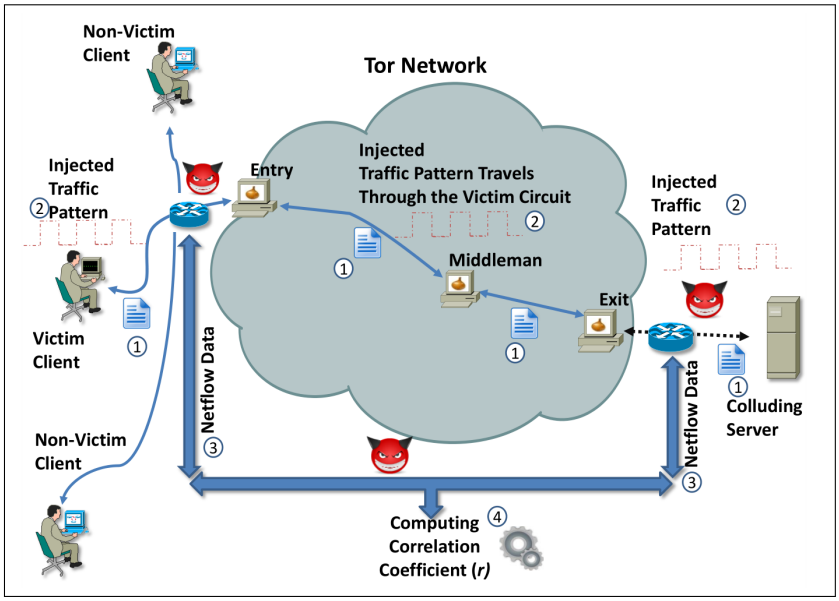


Abbildung 3.2: Angriff mittels Traffic-Analyse basierend auf NetFlow

Die Verfahren zur Deanonymisierung nach Murdoch und Chakravarty sehen einen mächtigen Angreifer vor, der Zugriff auf zentrale Knotenpunkte des Internets wie beispielsweise den DE-CIX hat. Weiterhin muss sich die Ressource auf der Seite der Austrittspunkte unter der Kontrolle des Angreifers befinden. Dadurch ist dieser in der Lage den angeforderten Datenstrom mittels *traffic-shaping* zu manipulieren.

3.1.3 Durchführung des Angriffs

Zur Bestimmung der optimalen Parameter für den Angriff wird das Analyseverfahren der NetFlow-Daten zuerst unter kontrollierten Laborbedingungen erprobt. Der Versuchsaufbau besteht aus 30 Teilnehmern (auf zwei Rechner verteilt), mehreren Tor-Relays und einem Server zu dem eine Verbindung durch das Tor-Netzwerk aufgebaut wird (Abbildung 3.3).

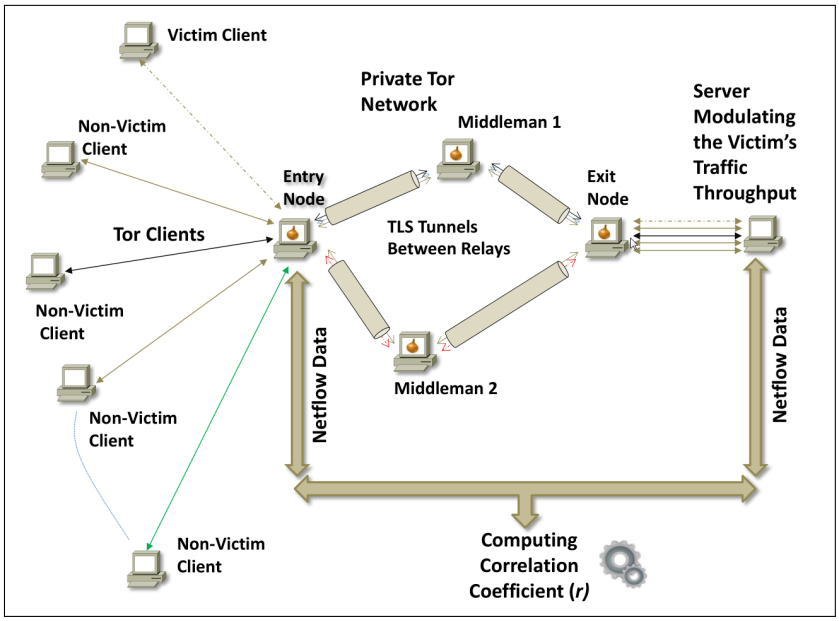
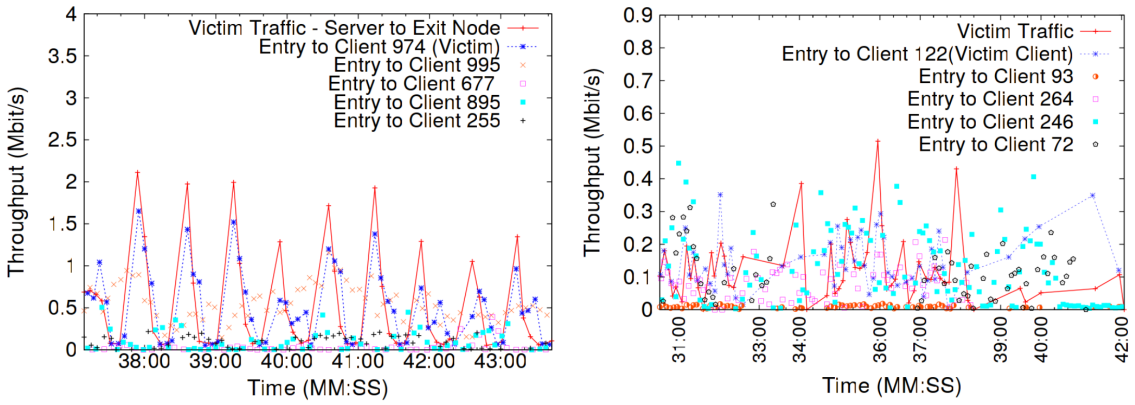


Abbildung 3.3: Aufbau eines privaten Tor Netzes unter Laborbedingungen

Weiterhin kommen OpenSource Werkzeuge *ipt_netflow* [15] und *flowtools* [16] zum Einsatz. Diese Werkzeuge haben gegenüber den Routern und Switches den Vorteil der einfacheren Parametrierung. Auch wenn ein Angreifer Zugriff auf NetFlow-Daten von Rechenzentren hat, ist er nicht zwangsläufig in der Lage, die Konfiguration zur Erstellung der NetFlows zu verändern. Sowohl die Daten der OpenSource Tools als auch die Flows der Router müssen nachträglich angeglichen werden, um eine zeitliche Übereinstimmung der aufgezeichneten Intervalle zu erreichen.

Der Angriff beschreibt ein Verfahren um bestimmte Nutzer gezielt aus der Masse der Verbindungen zu identifizieren. Das Opfer, welches es zu identifizieren gilt, muss von dem Server, der sich unter der Kontrolle des Angreifers befindet, einen zumindest kurzzeitigen, kontinuierlichen Datentransfer initiieren. Zur Veranschaulichung wird das Verfahren im Folgenden ausgehend von dem Download einer größeren Datei erläutert.



(a) Rechteckmuster (b) Komplexmuster

Abbildung 3.4: Korrelation von Client zu Server Flows. Abbildung mit den jeweils fünf höchsten Korrelationskoeffizienten

Der Angreifer verändert durch Bandbreitenmanipulation mittels *Linux Traffic Controler*[2] den Datenfluss der vom Opfer angeforderten Datei. Um die Auswirkungen auf den Erfolg der Flow Daten Korrelation zu überprüfen, wurden zwei verschiedene Muster benutzt (Abbildung 3.4).

Rechteckmuster Umschalten der Bandbreite des Clients alle 20 Sekunden von 2 Mbit/s auf 30 Kbit/s und zurück(Abbildung 3.4a)

Komplexmuster Umschalten der Bandbreite des Clients alle 20 Sekunden von 1 Mbit/s, 50 Kbit/s, 300 Kbit/s auf 100 Kbit/s und wieder von vorne beginnend (Abbildung 3.4b)

Im Idealfall lässt sich zwischen dem Muster, welches serverseitig erzeugt wird und dem Datenstrom vom Opfer zu seinem Eintrittspunkt in das Tor Netzwerk durch Korrelationsanalyse eine eindeutige Beziehung herstellen.

[2]http://www.lartc.org/

3.1.4 Korrelation der NetFlow Daten

Um eine Beziehung zwischen den einzelnen Flows herzustellen, wird eine statistische Korrelation nach Pearson [17] durchgeführt. Die Pearson-Korrelation ist ein Maß für den statistischen Zusammenhang zwischen zwei Datensätzen. Der empirische Korrelationskoeffizient liefert für eine Messreihe von gepaarten Ausprägungen (Flows) einen Wert von *-1* bis *+1*. Der Wert *0* deutet auf eine lineare Unabhängigkeit der Merkmale hin. Der Wert *+1* steht für positiven linearen Zusammenhang, hingegen *-1* für einen negativen.

Die Messungen des Laborversuchs (Kapitel 3.1.3) wurden insgesamt 60 mal wiederholt, wobei jede Messung über eine Dauer von 400 Sekunden durchgeführt wurde. Insgesamt wurden 30 Messreihen mit dem Rechteckmuster und 30 Messreihen mit dem komplexen Muster erstellt. Die Ergebnisse zeigen eine deutliche Korrelation der Flow-Daten vom Server zum Austrittspunkt und der Flows vom Opfer zum Eintrittspunkt (Tabelle 3.1).

Muster	**Korrelation Opfer**	**Korrelation Unbetroffene**
Rechteck	μ: 0.80, σ 0.08	μ: 0.06, σ: 0.16
Komplex	μ: 0.92, σ 0.06	μ: 0.07, σ: 0.14

Tabelle 3.1: Ergebnisse Laborversuch (μ: Korrelationskoeffizient, σ: Standardabweichung).

Die Ergebnisse des Laborversuchs ermöglichten eine Identifizierung des Opfers mit einer Erfolgsquote von 100% (Abbildung 3.5), was die Tauglichkeit der Traffic-Analyse unter Idealbedingungen bestätigt.

Zur Herstellung der Praxisrelevanz wurde der Versuch auf eine Analyse mit öffentlichen Tor-Knoten ausgeweitet. Der Eintrittsknoten sowie der Server befanden sich unter der Kontrolle des Angreifers in Spanien. Der Tor-Knoten bediente zur Zeit der Messung 1104 verschiedene Benutzer aus der ganzen Welt. Die Clients der Opfer, die es zu identifizieren gilt, wurden auf Texas in den Vereinigten Staaten, Leuven in Belgien und Korfu in Griechenland verteilt.

Für jeden Standort wurden die Messungen mit den beiden Mustern jeweils 15 mal wiederholt. Aus den 90 resultierenden Messreihen (Abbildung 3.6) konnten durch die Korrelation der Flow Daten 76 mal die Opfer korrekt identifiziert werden. Im Vergleich zu den Messreihen des Laborversuchs sind die Korrelationskoeffizienten deutlich geringer und liegen mit den Ergebnissen unbeteiligter Teilnehmer deutlich näher zusammen (Tabelle 3.2).

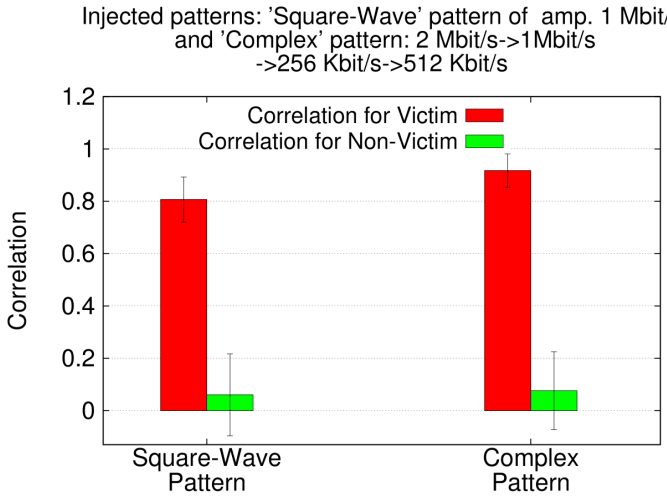


Abbildung 3.5: Ergebniss der Flow Korrelationen aus dem Laborversuch

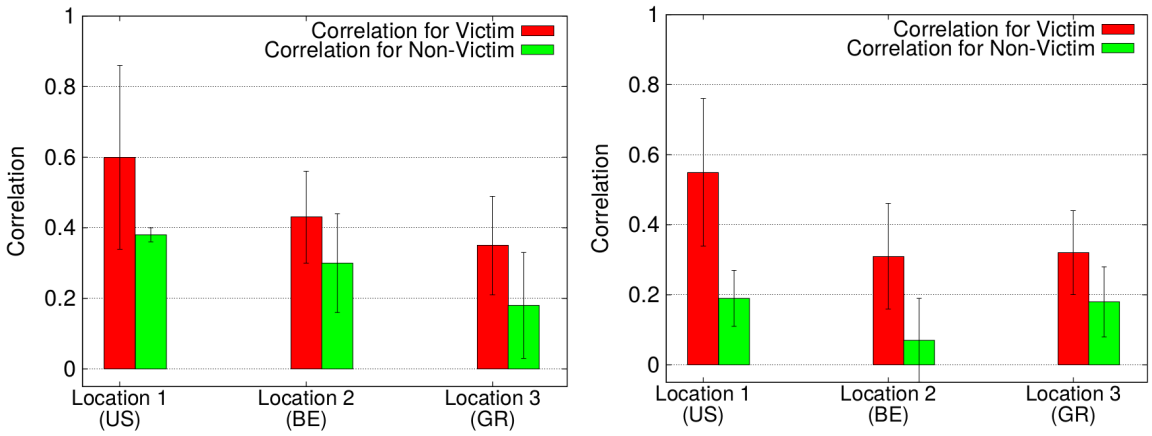
Standort	**Opfer (Rechteck/Komplex)**	**Unbetroffene (Rechteck/-Komplex)**
Texas (USA)	μ: 0.60, σ 0.26 / μ: 0.55, σ 0.21	μ 0.38, σ: 0.02 / μ 0.19, σ: 0.08
Leuven (Belgien)	μ: 0.43, σ 0.13 / μ: 0.31, σ 0.15	μ: 0.30, σ: 0.14 / μ 0.07, σ: 0.12
Corfu (Griechenland)	μ: 0.35, σ 0.14 / μ: 0.32, σ 0.12	μ 0.18, σ: 0.15 / μ 0.18, σ: 0.10

Tabelle 3.2: Maxima der Korrelationskoeffizienten für Opfer und unbeteiligte Teilnehmer. (μ: Korrelationskoeffizient, σ: Standardabweichung).

Die Ursache dafür ist unter Anderem die Verzerrung des injizierten Bandbreitenmusters durch das öffentliche Tor Netzwerk und der Empfindlichkeit des Pearson Korrelationskoeffizienten gegenüber Änderungen. Weiterhin befinden sich in den Ergebnissen für die öffentlichen Tor-Knoten auch falsch positive Treffer. Diese treten jedesmal auf wenn der Korrelationskoeffizient für einen unbeteiligten Teilnehmer höher als der des eigentlichen Opfers ist. Die Ursache liegt zum einen in der Überlastung des Tor-Netzwerks und dem Routing. Die Tor Relays versuchen die verfügbare Bandbreite auf alle Teilnehmer gleichmäßig aufzuteilen. Eine starke Manipulation des Datendurchsatzes eines Teilnehmers kann sich somit nach dem selben Muster auf andere Teilnehmer übertragen.

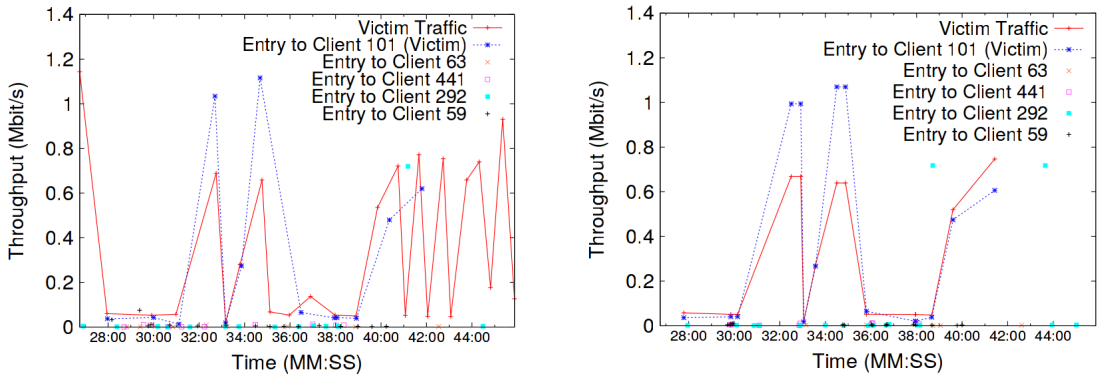
Die Ergebnisse basieren allesamt auf der Auswertung der Flows aus den OpenSource Werkzeugen *ipt_netflow* und *flowtools*. Zur Beurteilung des Analyseverfahrens wurden ebenfalls die Flow Daten des *edge routers*[3] des Labors ausgewertet. Durch die dort konfigurierten *active* und *inactive* Timer Einstellungen (60 und 15 Sekunden) musste

[3]Kanten-Router. Kante bzw. edge ist der Übergang zwischen zwei Netzen. Die Router werden nur an Netzübergängen eingesetzt und treten als Vermittler dieser in Erscheinung.



(a) Ergebnisse Rechteckmuster (b) Ergebnisse Komplexmuster

Abbildung 3.6: Durchschnittliche Ergebnisse der Korrelation mit öffentlichen Tor Relay für die drei verschiedenen Standorte der Opfer



(a) Diskrepanz der Messpunkte (b) Angleichung der Messpunkte

Abbildung 3.7: Gleichrichtung von Messpunkten für das Korrelationsverfahren

versucht werden die Flows des Servers mit denen des Routers anzugleichen, um eine zeitliche Übereinstimmung der Datensätze zu erreichen. Weiterhin entstehen durch unterschiedliche Flows Diskrepanzen in der Anzahl der korrelierbaren Datensätze, welche es zu normalisieren gilt (Abbildung 3.7). Aufgrund der besseren Ergebnisse der vorherigen Korrelationen durch das Komplexmuster wurde dieses für die folgenden Messungen ausgewählt. Die Intervallzeit nach der die Bandbreite geändert wird, wurde jedoch von 20 auf 30 Sekunden erhöht, um eine stärkere Auswirkung innerhalb des aktiven Flows zu erreichen.

Die Ergebnisse der Korrelation sind mit den angeglichenen Daten des vorherigen Versuchs vergleichbar. Von ebenfalls 90 Messungen konnten die Opfer in 71 Fällen korrekt identifiziert werden. Unter den Ergebnissen befanden sich sechs falsch positive Identifizierungen. In 13 Fällen war aufgrund des niedrigen Korrelationskoeffizienten ($< \mu\, 0.2$) keine Zuordnung möglich.

Eine erneute Ausweitung der Testparameter mit verschiedenen Tor Relays soll die Skalierbarkeit der Analyse zeigen. Die Anzahl der Teilnehmer während der Messungen lag mit zwei Tor Knoten bei 1500. Mit dem Komplexmuster wurden erneut acht Messungen für jeden der drei Standorte des Opfers durchgeführt. Von 24 Versuchen konnte das Opfer in 14 Fällen korrekt identifiziert werden, drei falsch positive und sieben nicht erkannte Teilnehmer. Die Gründe für eine Reduzierung der Messreihen und der abschließende Ausschluss der Ergebnisse in der Bewertung bleiben ungeklärt. Der Vollständigkeit halber werden im letzten Teil dieses Seminares alle Ergebnisse zur Abschlussbetrachtung herangezogen.

4 Zusammenfassung

Durch die verschiedenen Versuche wurde gezeigt, dass eine Deanonymisierung von Tor-Nutzer durch Traffic-Analyse grundsätzlich möglich ist. Die Auswertung der Korrelationsergebnisse zeigt jedoch einen immer weiter sinkenden Korrelationskoeffizienten, umso mehr der Versuchsaufbau sich einem realen Analyseszenario nähert (Tabelle 4.1).

Versuch	**Identifizierung**	**Falsch positiv**	**Falsch negativ**
Labor	60 / 60 (100 %)	0	0
Öffentlich (1 Relay)	76 / 90 (84.4 %)	4 / 90 (4.4 %)	20 / 90 (22.2 %)
Öffentlich (1 Relay, Flow durch Router vorgegeben)	71 / 90 (78.9 %)	6 / 90 (6.6 %)	23 / 90 (25.5 %)
Öffentlich (2 Relays, Flow durch Router vorgegeben)	14 / 24 (58.3 %)	3 / 24 (12.5 %)	7 / 24 (29.1 %)

Tabelle 4.1: Zusammenfassung der Ergebnisse aller Versuche

Besonders problematisch ist die erhöhte falsch Erkennung von 12.5 %. Während der Angreifer in der Lage ist die Flow-Daten des Opfers auf Seite des Servers exakt zu ermitteln, muss er auf Seite der Clients enorme Datenmengen analysieren (ausgehend von Flow Datensätzen die an einem Knoten wie dem DE-CIX anfallen). Bei 100.000 Flow Daten bedeutet eine falsch positive Erkennungsrate von 6 % eine unkorrekte Zuordnung bei 6.000 Nutzern. Bezogen auf ein reales Angriffsszenario wurde die Masse an in Frage kommenden Verbindungen zwar reduziert, jedoch ist eine korrekte Zuordnung eines Flows zu dem bestimmten Opfer nicht möglich [18].

Weiterhin haben die Versuche gezeigt, dass die Ergebnisse der Korrelation auch stark von der Auslastung und dem Bandbreitenmangement der einzelnen Tor Relays abhängig sind. In jedem Fall muss der zu identifizierende Client zu einem Datentransfer von einer Ressource bewegt werden die sich unter der Kontrolle des Angreifers befindet. Browsen im Web, Austausch von Chat-Nachrichten und Email-Verkehr erzeugen nicht genug Traffic um mit den beschriebenen Methoden manipuliert zu werden.

Abbildungsverzeichnis

2.1 Tor Netzwerk 4

3.1 NetFlow Datensatz 7
3.2 Angriff mittels Traffic-Analyse basierend auf NetFlow 8
3.3 Aufbau eines privaten Tor Netzes unter Laborbedingungen 9
3.4 Korrelation von Client zu Server Flows. Abbildung mit den jeweils fünf höchsten Korrelationskoeffizienten 10
3.5 Ergebniss der Flow Korrelationen aus dem Laborversuch 12
3.6 Durchschnittliche Ergebnisse der Korrelation mit öffentlichen Tor Relay für die drei verschiedenen Standorte der Opfer 13
3.7 Gleichrichtung von Messpunkten für das Korrelationsverfahren 13

Tabellenverzeichnis

3.1 Ergebnisse Laborversuch (μ: Korrelationskoeffizient, σ: Standardabweichung). 11
3.2 Maxima der Korrelationskoeffizienten für Opfer und unbeteiligte Teilnehmer. (μ: Korrelationskoeffizient, σ: Standardabweichung). 12

4.1 Zusammenfassung der Ergebnisse aller Versuche 15

Literaturverzeichnis

[1] M. Beckedahl, "Das sind die neuen pläne zur wiedereinführung der vorratsdatenspeicherung." https://netzpolitik.org/2015/das-sind-die-neuen-plaene-zur-wiedereinfuehrung-der-/vorratsdatenspeicherung/, Februar 2015.

[2] N. Sagener, "Entführt, erschossen, vergiftet." http://www.zeit.de/politik/ausland/2015-02/putin-moskau-nemzow, Februar 2015.

[3] "Bundesdatenschutzgesetz (bdsg)." http://www.gesetze-im-internet.de/bdsg_1990/__4.html, Februar 2015.

[4] "Telemediengesetz (tmg)." http://www.gesetze-im-internet.de/tmg/__13.html, Juli 2015.

[5] "Anonymität." https://de.wikipedia.org/wiki/Anonymität, September 2015.

[6] J. Michels, "Evaluierung von ausgewählten Anonymisierungsverfahren für das Internet," bachelor thesis, Fachhochschule Bonn-Rhein-Sieg, 2008.

[7] K. V. Dimitri Tokmetzis, "Metadaten: Wie dein unschuldiges smartphone fast dein ganzes leben an den geheimdienst übermittelt." https://netzpolitik.org/2014/metadaten-wie-dein-unschuldiges-smartphone-fast-dein-ganzes-leben/-an-den-geheimdienst-uebermittelt/, Juli 2014.

[8] The Tor Project, Inc, "About tor." https://www.torproject.org/about/overview.html.en, November 2015.

[9] J. Schmidt, "Eigen-Tor. Gefahren der Tor-Nutzung im Alltag," *c't*, vol. 20, p. 102, September 2013.

[10] S. J. Murdoch and P. Zieliński, "Sampled traffic analysis by internet-exchange-level adversaries," in *Proceedings of the 7th International Conference on Privacy Enhancing Technologies*, PET'07, (Berlin, Heidelberg), pp. 167–183, Springer-

Verlag, 2007.

[11] S. Chakravarty, M. V. Barbera, G. Portokalidis, M. Polychronakis, and A. D. Keromytis, "On the effectiveness of traffic analysis against anonymity networks using flow records." http://www.cs.columbia.edu/~sc2516/papers/pam2014-tor-nfattack.pdf, 2014.

[12] "Introduction to cisco ios netflow - a technical overview." http://www.cisco.com/c/en/us/products/collateral/ios-nx-os-software/ios-netflow/prod_white_paper0900aecd80406232.html, May 2012.

[13] "Netstream (integrated) technology white paper." http://enterprise.huawei.com/ilink/enenterprise/download/HW_201022, September 2012.

[14] "Inmon corporation's sflow: A method for monitoring traffic in switched and routed networks." https://www.ietf.org/rfc/rfc3176.txt, September 2001.

[15] "Netflow iptables module for linux kernel (official)." https://github.com/aabc/ipt-netflow, November 2015.

[16] "flow-tools - tool set for working with netflow data." https://code.google.com/p/flow-tools/, November 2015.

[17] Wanja Hemmerich, "Korrelation, korrelationskoeffizient." http://matheguru.com/stochastik/korrelation.html, 2015.

[18] Roger Dingledine, "Traffic correlation using netflows." https://blog.torproject.org/blog/traffic-correlation-using-netflows, November 2014.