Protección y Rompimiento del Pseudoanonimato de Blockchain

Víctor Reyes-Macedo*, Moisés Salinas-Rosales †
Instituto Politécnico Nacional
Centro de Investigación en Computación
Av. Juan de Dios Bátiz, Esq. Miguel Othón de Mendizábal
Nueva Industrial Vallejo, Gustavo A. Madero 07738
Ciudad de México
*vg.reyesmacedo@gmail.com,
†msalinasr@ipn.mx

Gina Gallegos-García
Instituto Politécnico Nacional
Escuela Superior de Ingeniería Mecánica y Eléctrica
Av. Santa Ana Nº 1000
San Francisco Culhuacan, Coyoacán 04260
Ciudad de México
ggallegos@ipn.mx

***Resumen*—A raíz del desarrollo de Bitcoin, la cadena de bloques o *blockchain*, que es la tecnología central de éste sistema de pagos, ha encontrado cada vez más espacios en los que su potencial puede ser aprovechado. Dichos espacios, han resultado particularmente atractivos a la industria, que ha adoptado esta tecnología en campos tan diversos como las finanzas, la medicina, la logística, el gobierno y la propiedad intelectual, por mencionar algunos. Sin embargo, aspectos tan importantes como la privacidad de las operaciones y el anonimato de sus participantes han sido dejados de lado. En este artículo, se presenta una revisión de los estudios que han evaluado propuestas para vencer el anonimato y la privacidad, y de los que han contribuido con propuestas para robustecer dichos servicios.**

***Index Terms*—Anonimato, bitcoin, blockchain, privacidad, seguridad .**

I. Introducción

En el año 2008, se presentó al mundo el sistema *Bitcoin*, un sistema de pago electrónico, el cual permite llevar a cabo transacciones de manera directa entre sus usuarios, sin la necesidad de involucrar a una entidad adicional de confianza, como es el caso de las instituciones bancarias. El funcionamiento de dicho sistema, en gran medida, se debe a la introducción de la cadena de bloques, conocida de manera global como *blockchain*, y que es descrita por Halpin y Piekarska en [1] como una lista de datos descentralizada y verificable criptográficamente, que garantiza la integridad de la información. Las aplicaciones de las cuales la tecnología *blockchain* forma parte al día de hoy, son numerosas y se presentan en diversos campos, siendo posible encontrar implementaciones en industrias dedicadas al cuidado de la salud, a las finanzas y servicios bancarios, elaboración de contratos inteligentes, licitaciones y servicios de gobierno, sistemas de votación electrónica y trazabilidad de insumos, entre otras [2]. El nivel y la velocidad de adopción del *blockchain* para una amplia gama de fines, pone de manifiesto la importancia del desarrollo de la investigación alrededor de la seguridad que ofrece este sistema, en particular, este documento aborda un enfoque centrado en los servicios de privacidad y anonimato. Por ello, este artículo presenta una revisión de las investigaciones que se han dedicado a estudiar estos aspectos, con la finalidad de ofrecer un panorama al respecto. El resto del artículo está organizado de la siguiente forma: la Sección II presenta un contexto del funcionamiento del blockchain de Bitcoin, la Sección III se centra en la revisión de estudios que han explotado el nivel de anonimato y privacidad del blockchain, mientras la sección IV presenta las propuestas de mejoramiento de estos servicios. La Sección V aborda los problemas abiertos al respecto, y en la Sección VI se presentan las conclusiones.

II. Una mirada a Bitcoin

Hablar de *blockchain*, frecuentemente conduce a hablar de Bitcoin para comprender la naturaleza de esta estructura de datos, cuya propuesta original, comprometió ligeramente la privacidad y el anonimato, con el objetivo de evitar el doble gasto y la falsificación durante las transacciones. Por ello, si bien las transacciones no se asocian, en primer instancia, a ninguna entidad, al completarse quedan registradas en el *blockchain* mediante identificadores, que consisten en cadenas alfanuméricas de entre 27 y 34 caracteres denominados *direcciones*. Así, una transacción consiste en el siguiente conjunto de datos:

- Dirección de origen: Son las direcciones que pagan un monto (puede ser más de una).
- Dirección de destino: Son las direcciones que reciben un monto (puede ser más de una).
- Monto: Cantidad de *bitcoins* que se pagan.
- Timestamp: Fecha y hora en que hizo la transacción.

Dichos datos, se almacenan en los bloques que conforman el *blockchain.* Estas características, demuestran que Bitcoin es un sistema de pago transparente, pese a que la identidad de los usuarios no es explícita. Por ello, diversas investigaciones se han centrado en los aspectos de privacidad y anonimato en el blockchain, con miras a fortalecer la seguridad en el rango de aplicaciones que tiene esta tecnología en la industria.

III. Rompiendo la privacidad y el anonimato

A menudo, los conceptos de privacidad y anonimato son confundidos, y pueden llegar a ser utilizados de manera indistinta. Al respecto, Bradbury señala en [3], que privacidad

significa ocultar el contexto, y anonimato significa ocultar al sujeto. En este sentido, al mejorar el servicio anonimato en *blockchain*, el objetivo será que el sujeto que interactúe en el sistema no sea identificable ni trazable, mientras que el servicio de privacidad deberá garantizar que la actividad de dicho sujeto no sea visible a terceros, es decir, la contraparte no debe tener acceso a los meta-datos de la interacción.

Al respecto, Kus Khalilov y Levi presentaron en [4], una taxonomía de los estudios de análisis de anonimato y privacidad en Bitcoin, la cual se muestra en la Figura 1.

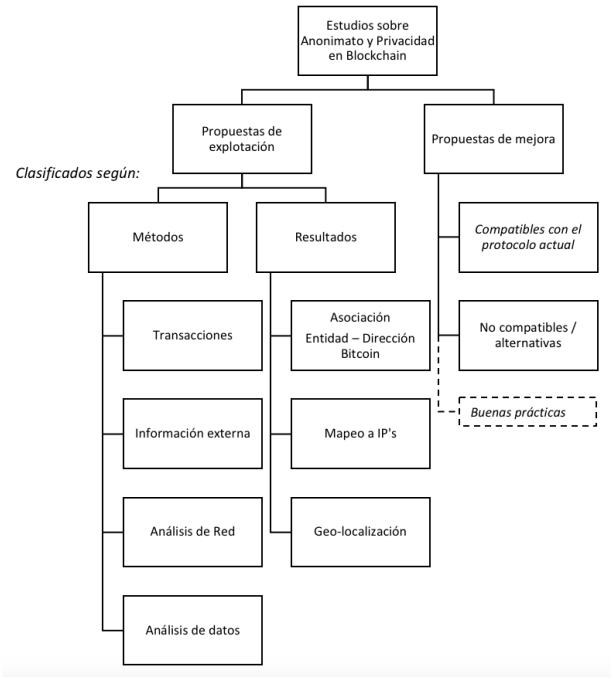

Figura 1. Taxonomía propuesta por Khalilov y Levi

Como se puede observar, proponen una clasificación con base en el tipo de resultados, la cual consiste en los siguientes aspectos:

- Identificación y asociación de una dirección Bitcoin con una persona o entidad, a partir de la información de la persona o entidad.
- Identificación y asociación de una identidad, a partir de la información de una dirección Bitcoin.
- Asociación de direcciones Bitcoin a direcciones IP.
- Asociación de direcciones Bitcoin que sean propiedad de un mismo usuario, mediante clústers.
- Geo-localización de direcciones Bitcoin.

En el mismo estudio, los autores abordan, además, los métodos que permiten llegar a los distintos tipos de resultados señalados arriba, a través de cuatro diferentes clases de métodos:

- Mediante transacciones
- Mediante uso de información externa
- Mediante el análisis del blockchain como red
- Mediante el análisis de los datos del blockchain

III-A. Mediante transacciones

Al interactuar con otro usuario, ya sea porque se compran o venden productos o servicios, necesariamente se conoce la dirección Bitcoin de la contraparte, por lo que ésta y su dirección pueden ser asociadas sin problema. Bajo este principio, Meiklejohn *et al.* [5] realizaron un *ataque de re-identificación*, en el cual llevaron a cabo operaciones con vendedores y proveedores de diversos servicios, con ello, lograron seguir el rastro de los pagos a través del *blockchain*, e identificar 344 transacciones con 87 entidades conocidas y 1070 direcciones.

Los autores, Kus Khalilov y Levi, incluyen en este apartado a los servicios de mixing o de anonimato, que tienen como objetivo mezclar los fondos de las transacciones de varias fuentes, y luego hacer los pagos correspondientes, para así hacer que dichos pagos no sean rastreables. Un análisis de este tipo de servicios y el nivel de anonimato que ofrecen, puede ser consultado en el trabajo de Möser *et al.* en [6].

III-B. Mediante el uso de información externa

En muchos casos, de manera voluntaria o involuntaria, los usuarios suelen revelar información acerca de la propiedad de sus direcciones Bitcoin, por ejemplo, para recibir un pago o donación. También es usual que en foros en línea, especializados o no en el tema, se publiquen direcciones que tienen altos índices de actividad -grandes montos durante amplios períodos, como el caso de los servicios de cambio de criptomonedas o *exchange*- y se asocien con las entidades correspondientes. Mediante este método, Reid y Harrigan identificaron algunas entidades asociadas a un presunto robo de 25000 BTC; además propusieron la heurística de *transacciones multi-entrada* para identificar a los propietarios de determinadas direcciones [7]. De la misma forma, Ron y Shamir asociaron 1088 transacciones de 83 direcciones con WikiLeaks, a partir de la dirección Bitcoin que la organización utiliza para recibir donaciones [8]. Por otra parte, Ortega desarrolló scripts para ligar direcciones Bitcoin con la identidad de sus usuarios, a través de la información que éstos proporcionaron en foros públicos [9].

Spagnuolo, posteriormente, presentó en [10] el *framework* BitIodine, el cual es un analizador del *blockchain*, que tiene la capacidad de agrupar las direcciones que son propiedad de la misma entidad en clusters, y etiquetarlos. Baumann *et al.* [11], lograron obtener las direcciones IP asociadas con las direcciones Bitcoin a través del sitio `blockchain.info`, con lo cual identificaron un conjunto de direcciones pertenecientes a MtGox, una antigua empresa operadora de criptomonedas. Finalmente, Lischke y Fabian reunieron información sobre 223,000 direcciones IP que fueron usadas en 15.8 millones de transacciones [12].

III-C. Mediante el análisis del blockchain *como red*

Este método utiliza la información de las transacciones, obtenida a través del análisis del tráfico en la red de Bitcoin o su infrastructura de red. Para ello, los enfoques que los

autores del estudio [4] identificaron, fueron clasificados en las siguientes categorías:

- *Utilizando transacciones retransmitidas de manera anormal*: Se definen patrones de comportamiento anormal, Koshy *et al.* en [13] utilizaron este método para relacionar direcciones Bitcoin con direcciones IP.
- *Utilizando el primer retransmisor de información*: Consiste en suponer que el primer nodo en informar de una transacción es el origen de ésta. No obstante, resulta poco efectivo comparado en el método anterior.
- *Utilizando nodos de entrada*: Esta información puede ser obtenida por los nodos al conectarse a la red, y de esta forma es posible identificar a los propietarios de las transacciones y relacionar las direcciones con sus respectivas IP's.
- *Estableciendo una dirección cookie para la huella del usuario*: permite asociar direcciones IP y direcciones Bitcoin del mismo usuario, de manera que es posible obtener una huella del propietario. El método fue propuesto por Biryukov y Pustogarov en [14].

III-D. Mediante el análisis de los datos del blockchain

Dado que la información y los metadatos de todas las transacciones generadas son de acceso público en el *blockchain*, el flujo de bitcoins es trazable. Dicha información puede ser consultada en sitios especializados como `blockchain.info`, o bien, para obtener una copia es suficiente con descargar algún cliente Bitcoin, como BitcoinCore, el cliente original. De esta forma, el estudio de Reid y Harrigan, fue el primero que abordó la privacidad y el anonimato en el *blockchain* de Bitcoin. Con la información disponible, representaron el flujo de pagos como una red dirigida, en la cual las direcciones tomaron el rol de nodos, y los enlaces indicaban los montos de las transacciones. Posteriormente, plantean que las direcciones que son propiedad del mismo usuario pueden ser identificadas mediante tres heurísticas: a) transacciones multi-entrada, b) direcciones de cambio y c) clustering basado en comportamiento

Androulaki *et al.* retomaron las heurísticas anteriores y llevaron a cabo una medición del anonimato, mediante una simulación del entorno Bitcoin. Ron y Shamir, por otro lado, aplicaron estas heurísticas para asociar las transacciones y direcciones relacionadas con *The Sheep Market Place*, un mercado negro en *deep web*, y al caso de *Dread Pirate Roberts*, supuesto creador del mercado negro conocido como *Silk Road*. Ortega *et al.* realizaron un análisis con cada heurística, con el cual demostraron que la forma más eficiente de asociar direcciones es mediante *transacciones multi-entrada*. Baumann *et al.* y, de forma independiente, Lischke y Fabián, llevaron a cabo un proceso similar. Meiklejohn *et al.* utilizaron las dos primeras heurísticas para ejecutar un ataque de reidentificación mediante información externa, mientras Spagnuolo desarrolló *BitIodine*, un framework que logra asociar y etiquetar direcciones en clusters, dicho framework tomó como caso de estudio las transacciones asociadas con *Dread Pirate Roberts* y el ataque del ransomware *CryptoLocker*. Mediante la aplicación de estas mismas heurísticas, Ober *et al.* descubrieron que el anonimato en Bitcoin se reduce a medida que el tamaño de las entidades que participan y el comportamiento de las operaciones se vuelven estacionarios. En el caso del estudio realizado por Dupont y Squicciarini, lograron geolocalizar a las entidades identificadas. Ferrin, al aplicar estas heurísticas, sugirió una forma de diferenciar a las direcciones de cambio, y Yanovich *et al.* reportaron que aproximadamente un 2.5 % de las transacciones pasan por un servicio de *mixing* para asegurar su anonimato. Por otra parte, Nick introdujo dos heurísticas nuevas, d) heurística de consumidor y e) heurística de cambio óptimo, que posteriormente fueron retomadas por Neudecker y Hartenstein [4].

Autor	Inicial (α)	Final (β)	Reducción ϕ
Reid y Harrigan	1,253,054	881, 678	0.7036
Androulaki *et al.*	1, 632,648	1, 069, 699	0.6552
Ron y Shamir	3,730, 218	2,460, 814	0.6597
Ortega*	1, 719, 312	32,956	0.0191
Ortega **	383, 990	32, 261	0.0840
Baumann*et al.*	17, 229, 680	No especificado	No especificado
Lischke y Fabián	17,229,680	No especificado	No especificado
Meiklejohnn	12, 056, 684	3, 383, 904	0.2806
Spagnuolo	18,153, 279	3, 383, 904	0.1864
Möser	No especificado	No especificado	No especificado
Ober *et al.*	12,711,115	No especificado	No especificado
Dupont y Squicciarini	38,886,789	17,472,156	2.225
Zhao y Guan	35,770,360	13, 062, 822	2.738
Fleder *et al.*	80, 030	54,941	1.456
Ferrin	112,070,000	No especificado	0.7036
Yanovich *et al.*	No especificado	No especificado	0.7036
Neudecker y Hartenstein	196,963, 722	$\sim$ 72 millones	02.735
Nick	60,880,000	No reportado	0.3

*Mediante análisis de transacciones con multientrada
**Mediante análisis de direcciones de cambio

Tabla I
MEDICIÓN DEL PARÁMETRO DE *reducción* DE LOS ESTUDIOS

Finalmente, Kus Khalilov y Levi, en [4], realizaron un análisis comparativo de los diferentes estudios que se han hecho al respecto, con énfasis en los métodos seguidos y los resultados obtenidos. La tabla I muestra una medición de la eficiencia de los estudios mencionados, al asociar direcciones con entidades. En dicha tabla, la primer columna señala al autor o autores del estudio en cuestión, la segunda, indica el número de direcciones con el cual inició el estudio y la tercera columna muestra el número de grupos de direcciones obtenidos. Para medir la eficiencia en el proceso de formación de grupos de direcciones, se propone el parámetro de *reducción* ϕ, definido por la ecuación 1.

$$\phi = \frac{\alpha}{\beta} \tag{1}$$

Donde α es el número de direcciones inicial, y β es el número de direcciones final, dado por el número de grupos

obtenidos. Este parámetro, no contempla las ventanas de tiempo correspondientes a la obtención de las muestras de direcciones, e indica que el estudio correspondiente tuvo un mejor desempeño respecto a otro si el valor es más cercano a cero. Este resultado se indica en la última columna de la tabla I. La leyenda *No especificado*, se debe a que el estudio no reporta la cantidad de direcciones con las que trabajó.

IV. Protegiendo la privacidad y el anonimato

Derivado de los estudios mencionados anteriormente, se han dado a conocer algunas medidas que tienen la capacidad de mejorar el nivel de privacidad que el sistema Bitcoin provee. En el estudio de Reid y Harrigan [7], los autores proporcionan las siguientes recomendaciones: a) generar nuevas direcciones para cada transacción, b) evitar revelar información que facilite la asociación de la dirección con la identidad del usuario, c) enviar fracciones de Bitcoin a una dirección propia de reciente creación y d) usar un servico de mixing confiable. Adicionalmente, Androulaki sugirió en [15], dividir el monto que se requiera para hacer un pago a una dirección propia, con el objetivo de evadir la heurística de direcciones de cambio. Por otro lado, Biryukov *et al.* [14] sugiere incrementar el tiempo requerido de computación para cada conexión, así como agregar retrasos aleatorios antes de completar las transacciones, para elevar la dificultad de la asociación de direcciones. Sin embargo se considera poco viable debido a que afectaría de manera negativa la funcionalidad del proceso. Ortega, además, propone en [9] el uso de diferentes monederos para distintos propósitos, de esta manera podría evitar la asociación de direcciones, y adicionalmente, incluir un pago lo suficientemente pequeño, con varios números en la parte fraccionaria, para complicar la asociación de direcciones de origen y destino que podría generar la heurística de dirección de cambio.

Más allá de las recomendaciones generales, que pueden ser consideradas como *buenas prácticas*, existe una variedad de propuestas que buscan mejorar los aspectos de seguridad que tienen que ver con anonimato y privacidad. Kus Khalilov y Albert Levi presentaron en [4], una clasificación de los estudios que ofrecen propuestas que pretenden proteger las operaciones en *blockchain* contra los intentos de identificación.

Dicha clasificación comprende dos ramificaciones principales: la primera, engloba los procesos que son compatibles con el protocolo actual de Bitcoin, sin la necesidad de implementar modificaciones de algún tipo, a los cuales denomina *backwards compatible*. Por otra parte, la segunda ramificación refiere a los procesos que no son compatibles con el protocolo actual, y para los cuales, sería necesario proponer alternativas a Bitcoin o bien, modificar su actual implementación, a esta rama se le denominó *not backwards compatible*.

Los procesos *backwards compatible*, básicamente consisten en alguna forma de mixing, técnica mediante la cual se cruzan los pagos de transacciones, para que no lleguen de forma directa [16] y que puede darse de forma centralizada o descentralizada. Mientras tanto, los métodos *not backwards compatible* se enfocan en: mezcla de direcciones ocultas, mezcla de propiedad, cifrado de datos y desintegración de datos.

Así mismo, presenta una clasificación en sub-categorías, tomando como criterios el enfoque, los protocolos y los métodos usados en cada estudio; además, ubican los resultados en una o varias de las siguientes categorías: a) rompimiento de relación entre direcciones de origen y destino en una transacción, b) rompimiento de relación entre transacciones, c) ocultamiento de montos de pago y d) ocultamiento de direcciones IP [4].

Finalmente, Andrew Miller presentó en [17], un framework que preserva la privacidad de los contratos inteligentes, denominado *Hawk*, con la intención de que cualquier programador sea capaz de programar un contrato inteligente sin tener que implementar funciones criptográficas, de manera que el compilador, automáticamente, compila el programa a un protocolo criptográfico entre los usuarios y el *blockchain*.

V. Problemas Abiertos

Kus Khalilov y Albert Levi consideran, en [4], que aunque existan propuestas de mejora al anonimato y a la privacidad, la expansión que experimentan los sistemas basados en Blockchain llevarán a mayores avances en los campos de la criptografía y la computación sobre los cuales habrá que dirigir los esfuerzos de investigación, principalmente en cuatro aspectos:

- *Desempeño:* Los métodos desarrollados para mejorar el nivel de privacidad y anonimato, deben incluir una sólida investigación respecto al desarrollo de métodos más efectivos, computacionalmente.
- *Seguridad:* Las propuestas criptográficas en torno al incremento de los aspectos de seguridad y privacidad deben ser examinados en busca de posibles vulnerabilidades.
- *Escalabilidad:* Un reto no menor, es conseguir las mejoras necesarias a los protocolos, sin comprometer la escalabilidad del sistema.
- *Anonimato y confianza:* Se debe buscar la forma de balancear el anonimato y la confianza, debido a que, a mayor nivel de anonimato, el nivel de confianza en el sistema es susceptible de disminuir, la razón de ello es la limitada capacidad que tendrían los usuarios de verificar el buen funcionamiento del mismo.

VI. Conclusiones

Como se ha podido observar, la idea original del *blockchain* implementada en Bitcoin, en el mejor de los casos provee pseudo-anonimato, contra el cual diversos estudios han dirigido esfuerzos para vencer, logrando en muchos casos identificar a las partes involucradas en las transacciones, mediante una variedad de métodos. Por otro lado, también existen estudios que se han dedicado a fortalecer el nivel de privacidad que dicha implementación ofrece, obteniendo dos tipos de resultados: aquellos que pueden implementarse sin necesidad de modificar el protocolo actual, y aquellos que, por el contrario, se plantean como una alternativa ya que proponen una modificación a dicho protocolo. No obstante, si bien éstas últimas logran su objetivo respecto a las garantías de seguridad, lo hacen

a cambio de poder de cómputo y riesgos a la integridad de la información. Adicionalmente, un reto interesante está en el desarrollo de métodos que permitan cuantificar y comparar el anonimato y privacidad obtenidos de los diferentes estudios, con la finalidad de diferenciar aquellos que logran mejores resultados. De esta forma, con miras a establecer protocolos que permitan implementar *blockchain* para detonar todo su potencial en la industria, es importante dirigir esfuerzos que concilien los aspectos de desempeño, seguridad, escalabilidad y confianza.

REFERENCIAS

[1] Harry Halpin and Marta Piekarska. Introduction to security and privacy on the blockchain. In *Security and Privacy Workshops (EuroS&PW), 2017 IEEE European Symposium on*, pages 1–3. IEEE, 2017.

[2] Michael Crosby, Pradan Pattanayak, Sanjeev Verma, and Vignesh Kalyanaraman. Blockchain technology: Beyond bitcoin. *Applied Innovation*, 2:6–10, 2016.

[3] Danny Bradbury. Anonymity and privacy: a guide for the perplexed. *Network Security*, 2014(10):10–14, 2014.

[4] Merve Can Kus Khalilov and Albert Levi. A survey on anonymity and privacy in bitcoin-like digital cash systems. *IEEE Communications Surveys & Tutorials*, 2018.

[5] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140. ACM, 2013.

[6] Malte Moser, Rainer Bohme, and Dominic Breuker. An inquiry into money laundering tools in the bitcoin ecosystem. In *eCrime Researchers Summit (eCRS), 2013*, pages 1–14. IEEE, 2013.

[7] Fergal Reid and Martin Harrigan. An analysis of anonymity in the bitcoin system. In *Security and privacy in social networks*, pages 197–223. Springer, 2013.

[8] Dorit Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph. In *International Conference on Financial Cryptography and Data Security*, pages 6–24. Springer, 2013.

[9] Marc Santamaria Ortega. The bitcoin transaction graph anonymity. 2013.

[10] Michele Spagnuolo, Federico Maggi, and Stefano Zanero. Bitiodine: Extracting intelligence from the bitcoin network. In *International Conference on Financial Cryptography and Data Security*, pages 457–468. Springer, 2014.

[11] Annika Baumann, Benjamin Fabian, and Matthias Lischke. Exploring the bitcoin network. In *WEBIST (1)*, pages 369–374, 2014.

[12] Matthias Lischke and Benjamin Fabian. Analyzing the bitcoin network: The first four years. *Future Internet*, 8(1):7, 2016.

[13] Philip Koshy, Diana Koshy, and Patrick McDaniel. An analysis of anonymity in bitcoin using p2p network traffic. In *International Conference on Financial Cryptography and Data Security*, pages 469–485. Springer, 2014.

[14] Alex Biryukov and Ivan Pustogarov. Bitcoin over tor isn't a good idea. In *Security and Privacy (SP), 2015 IEEE Symposium on*, pages 122–134. IEEE, 2015.

[15] Elli Androulaki, Ghassan O Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. Evaluating user privacy in bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 34–51. Springer, 2013.

[16] Malte Moser. Anonymity of bitcoin transactions. 2013.

[17] Andrew Miller. *Provable security for cryptocurrencies*. PhD thesis, 2016.