Weighted Factors for Evaluating Anonymity

Khalid Shahbar(✉) and A. Nur Zincir-Heywood

Dalhousie University, Halifax, Canada
{Shahbar,Zincir}@cs.dal.ca

Abstract. Many systems provide anonymity for their users, and most of these systems work on the separation between the users' identity and the final destination. The level of anonymity these services provide is affected by several factors, some of which are related to the design of the anonymity service itself. Others are related to how the system is used or the user's application/purpose in using the anonymity service. In this paper we: (i) propose five factors that aim to measure anonymity level from the user's perspective; (ii) evaluate these factors for three anonymity services, namely Tor, JonDonym, and I2P as case studies; and (iii) present a mechanism to evaluate anonymity services based on the proposed factors and measure their levels of anonymity.

Keywords: Anonymity factors · Metrics · Tor · Jondonym · I2P

1 Introduction

There are many tools, applications, and websites on the Internet claiming to protect the privacy of their users. The levels of privacy protection provided by these services are different based on the way they work. For example, VPN (Virtual Private Network), which can be provided either as a free or a paid service, hides the user's identity while surfing the Internet anonymously. However, the VPN service provider has access to the user's identity and his/her activity on the Internet. Some of these service providers also keep the logs of their users. This is also the case with free proxy websites, which claim that they protect the user's identity.

Tor, JonDonym, and I2P are popular anonymity services. They provide anonymity to their users to hide their identity from Internet web servers and hide the websites they have accessed. These systems prevent not only the web servers from revealing users' identities, but also the operators of the systems themselves from identifying the users. However, there are many details behind this kind of anonymity that might not be clear or obvious to the user.

Therefore, the anonymity level of the users is not the same, even when using an anonymizing tool. The reason behind using an anonymity service varies from one user to another. This could affect the anonymity level and the choice of the

A. Imine et al. (Eds.): FPS 2017, LNCS 10723, pp. 303–318, 2018.
https://doi.org/10.1007/978-3-319-75650-9_20

right anonymity service. The design of the anonymity tools varies based on: (i) Which services such a tool offers to users, and (ii) How the user decides or measures anonymity level, given all the different anonymity services. In this paper, we present a method of calculating and comparing user anonymity levels that takes into consideration the different needs of different users to answer the aforementioned questions. Therefore, we aim to assist the user in choosing the most suitable anonymity service for their needs. The proposed method depends on evaluating anonymity systems based on five factors. To measure the anonymity level using this method, the factors are converted to numeric values in order to assign weights and scores. In addition, each factor is compared with the others according to the goal or purpose of anonymity. Therefore, the relative weights (importance) of the factors are determined based on who is using the anonymity service and why. In doing so, our objective is to provide a comprehensive measurement technique that could be used to evaluate the level of anonymity based on the environment in which the anonymity service is used.

The rest of this paper is organized as follows. The related literature is reviewed in Sect. 2. The Tor network, the JonDonym network, and the I2P network are discussed in Sect. 3. Section 4 presents and discusses the five factors regarding the level of privacy in anonymity services studied in this work, and Sect. 5 evaluates these anonymity factors. Finally, conclusions are drawn and potential future work is discussed in Sect. 6.

2 Related Literature

Measuring the anonymity level is a challenge for a number of reasons. One is the difference in the design and the goal of the anonymity systems (networks). On the other hand, there is no single way to measure anonymity levels on different anonymity networks. In addition, anonymity level is not directly quantifiable as compared to other network traffic measurements such as delay, bandwidth, volume, etc. In [13], Ries et al. evaluated five anonymization tools with regard to performance, usability, anonymity, network reliability, and cost. The evaluated tools were Tor, I2P, JonDonym, Perfect Privacy and Free proxies. Performance factors used to evaluate and rank these tools were Round Trip Time (RTT), Inter-Packet Delay Variation (IPDV), and throughput. Additionally, they used installation, configuration, and verification of the anonymization connection as factors to define the usability of these tools.

Dhiah el Diehn et al. examined the usability of four anonymity tools (Tor, JonDo, I2P, and Quicksilver) during the installation phase [1]. They detailed the installation process of these tools, applying four tasks to test the installation phase: success of installation, success of configuration, confirmation of anonymization, and ability to disable anonymization. To test the usability of these tools, they used eight guidelines from [3], which focused on the user's ability to perform the four tasks mentioned above.

Wendolsky et al. compared Tor and AN.ON (JonDonym) from the user's perspective, based on performance and number of users [18]. Latency and bandwidth

were used to measure performance, and the results showed that Tor performs unpredictably based on time of day.

The above studies focused mainly on evaluating anonymity services based on their performance or usability, where anonymity was not the focus of the evaluation. On the other hand, there are studies where measuring anonymity was the main goal. In these cases, the idea of measuring anonymity is about minimizing the ability of an attacker to correlate the sender and the receiver, even if they communicate over a channel observed by the attacker. To anonymize against such a threat model, Chaum [2] presented the concept of the "anonymity set", in which the set is the total number of participants in the anonymity service that may include the sender. When the size of the set is increased, the anonymity level is considered to increase as well.

Serjantov and Danezis [14] developed the concept of the anonymity set by using the information-theoretic metric based on anonymity probability distribution. Diaz et al. [4] also used an information-theoretic model to evaluate the anonymity level of a system in a particular attack scenario.

Murdoch [12] surveyed studies performed on measuring anonymity for low-latency anonymous networks and high-latency email anonymous networks and discussed the development of the techniques used for measuring anonymity.

Even though the above studies have been important in measuring anonymity levels, the "anonymity" of the anonymity services is affected by other factors, too, such as the users' behaviors and browsers settings. Therefore, in this research, we present a method to measure the level of anonymity by analyzing the anonymity service from different perspectives and propose metrics (factors) that enable us to measure the anonymity of such services. The anonymity set which is presented by Chaum [2] is a way to measure the level of anonymity on the multilayer-encryption anonymity networks. It presents the number of possible choices to which a message on the anonymity network belongs for a specific user. The higher the value of the anonymity set, the better the anonymity becomes. This way of measuring the anonymity level focusses on the probability of linking the message to the user. However, the level of anonymity could be affected by other factors, too. Therefore, in this research, weighted factors for measuring anonymity services are presented as another way to measure and quantify the anonymity level. The method takes into consideration multiple factors: quantifying, comparing and applying them to evaluate the anonymity level of different use cases.

3 Anonymity Systems Studied

Multilayer-encryption anonymity networks share the goal of providing anonymous services to their users. The anonymity services vary in terms of design, performance, delay, and provided services. The following introduces the most popular multilayer-encryption anonymity networks: Tor, JonDoNym, and I2P.

3.1 Tor Network

The Tor network is based on volunteers to run their machines as Tor relays (also called routers or nodes). Tor provides anonymity to its users by hiding the IP addresses of the users and by hiding the content of the users' traffic, as long as that traffic is still on the Tor network. The IP addresses of the users are hidden by relaying all the users' requests through the Tor network. The users' traffic is hidden by dividing the packets into smaller fixed-size encrypted cells. Tor also provides a service called Hidden Services that hides the IP address of a web server for users who want to keep their identities hidden.

There are three types of nodes on the Tor network: entry node, middle node and exit node. The entry node is the first node that the user communicates with when trying to establish a circuit to carry their traffic. The middle node is an intermediate node that lies between the entry node and the exit node, and the exit node is the node used to relay the user's request to the web server. Since all three types of nodes are run by volunteers, running an exit node is an optional choice available while configuring the node to run in the Tor network. The exit node has the option to be configured for allowing certain types of traffic based on the port number. This enables the volunteered user who runs the exit node to determine the type of traffic to block/pass through the exit node.

Whenever the user sends his/her traffic through Tor, a virtual circuit is used to relay the user's traffic. The virtual circuit consists of a connection of the three types of nodes (entry, middle, and exit nodes). The user starts by establishing a TLS (Transport Layer Security) connection with the first node. After the connection is made with the first node, the user requests that the entry node extend the connection to the middle node. Finally, the connection is extended again to the exit node. The Tor browser is responsible for translating all the user's requests to the virtual circuit. This includes hiding the IP address of the user, dividing the packets into smaller cell(s), encrypting the traffic with three layers of encryption, receiving the data from the web server that comes in encrypted cells and decrypting the received cells.

3.2 JonDoNym

JonDonym is a network of mix cascades, providing anonymity to the users based on multilayer encryption. The cascade consists of two (free) or three (paid) mix servers. The user starts the connection to the JonDonym network by selecting the mix cascade. Currently there are five free cascades and eleven paid cascades the user can choose from.

Only one active connection to one cascade is possible during the user's connection to the JonDonym network. Each HTTP request will create a connection from the browser (JonDoFox) with the client software JonDo. The JonDoFox browser can generate multiple connections with the JonDo. All these connections are multiplexed into one connection to the first mix server, which receives connections from multiple users. All the users' connections are then multiplexed

into one TCP/IP connection to the second mix, or to the last in case of only two mixes in the cascade.

The information about the available cascades, the number of users, the loads, and the mix status are stored in the InfoService [9]. The user gets the information about the cascades from the InfoService, and the last mix sends the users' requests to cache proxies. Multilayered encryption is used during the communication between the user and the last mix, which ensures that even the mixes cannot access the user's data. The path that the user's data takes is fixed based on the chosen cascade. To choose another path (cascade), the user has to start a new connection to the JonDonym. The user can only have one connection to one cascade at any given time.

3.3 I2P

I2P network is a decentralized anonymous network with no central database or server that contains the network database. The network database (netDb) is distributed by using the Kademlia algorithm [10], which is used in many applications where peer-to-peer (P2P) communication is needed in a decentralized network. The information that the user gets from the netDb enables the user to build tunnels. Communications over I2P require inbound and outbound tunnels, which are unidirectional. The netDb contains the leaseSet of the tunnels and routers. LeaseSet shows the routers involved in a tunnel. RouterInfo in the netDb shows how to contact a specific router. The user has the option to modify the number of routers in the outbound tunnel. I2P uses the concept of garlic routing [5], where layered encryption is implemented in addition to binding multiple messages together. The messages within the I2P network are encrypted end-to-end as long as the two communication parties are within the I2P network. However, when the user communicates with an end-system that is outside of the I2P network using an outproxy, then the encryption is not end-to-end.

By default, the user within the network transfers their data and that of other users where the user's machine functions as a resource for the network. The user can change the amount of bandwidth dedicated to the network from the console. The users' contributions in relaying the network data are restricted by relaying the data only within the I2P network. A different configuration is required when a user wants to relay the I2P traffic to an end-system outside of the I2P network (outproxy). The number of outproxies in the I2P network is limited.

One of the major differences between I2P and other anonymity networks such as Tor and JonDonym is that I2P is designed as a private network. The users mainly communicate within the network. The user builds two tunnels: inbound and outbound. The inbound tunnels are used to receive messages, and the outbound tunnels are used to send messages.

4 Proposed Factors

This section presents the five proposed factors to analyze the anonymity level of the aforementioned anonymity systems (Tor, JonDonym, and I2P).

4.1 The Level of Information Available for the Service Provider

When a Tor user connects to the Tor network, a virtual circuit is created. The circuit consists of three nodes; the first node has the actual IP address of the user, and therefore his identity, but it does not have knowledge of his Internet activity. This information could be used to perform attacks that depend on the correlation between the duration, data, and the server. The exit nodes, through which all the requests of the users are relayed, have a considerable amount of information, as they are the links between the Internet and the Tor users. The operator of the exit node has the ability to know and statistically evaluate the user's activities on the Tor network [11]. Another important fact about the exit node that might not be clear for non-technical users is that the encryption of the requests through the exit node are all based on the encryption of the original requests and has nothing to do with the three levels of encryption on the Tor Network. Therefore, the exit node alone can breach the anonymity of the users if they use their login information to access their email or any web server without sending an encrypted request.

Furthermore, JonDonym works in a similar fashion to Tor. The first point on the JonDonym network (First Mix) receives the connection request from the user, which has the information about the connection duration and the user's identity. The last point (Last Mix) does not know the user's identity, but it has the activities or the websites that the users request. The encryption layers used by JonDonym and Tor overlay networks protect the data, even from the operators; an exception is when the data sent by the user to the webserver is not encrypted; then, the last node/mix has the ability to access the data sent by the user. The anonymity mechanism in Tor/JonDonym depends on relaying the user data through multiple points (Node/Mix). Each node/mix only knows part of the connection information, not the whole information required to connect the user to the request.

On the other hand, what if all the nodes/mixes on the path between the user and the server are compromised or attacked? On the Tor network, the three nodes in the circuit path are selected by using the path selection protocol, which specifies the three nodes the user will use to relay the data in conjunction with the policy that the exit node operator defines. In addition, the user has the ability to override the path selection protocol and to choose a specific exit/entry node. This flexibility and randomness in node selection makes it harder for an attacker to target a specific user by trying to compromise the three nodes that the user selects. Moreover, running and compromising three nodes do not mean that these three nodes will be selected by the path selection protocol.

On the JonDonym network, this type of attack is also possible; the difference lies in the operation of the mixes. The number of mixes on the JonDonym is fewer than the number of nodes on the Tor network. But, the operators of the mixes are registered with their identities. They also sign an agreement with JonDonym not to exchange information between operators of the mixes and not to save user data. One of the differences between Tor and JonDonym is that JonDonym mixes do not change, and the path is always the same. In the case

of cooperation between all the mixes, it is possible to breach user anonymity on the JonDonym network.

Last but not the least, the goal of the I2P network is different than Tor and JonDonym. I2P is designed to provide anonymity for the users within the I2P network; However, that does not mean that I2P services are limited by the network boundaries. Browsing webpages outside the I2P network requires configuring the user's machine to use an outproxy. In this case, the information available to the outproxy is similar to Tor's exit router or JonDonym's last mix. The outproxy has access to all traffic passing through; if the traffic is not encrypted, the outproxy can see sensitive information.

The common point between the three anonymity services is that at any point, part of the network has some user information. This information could be the IP address of the user that is available at the first point on the anonymity network that connects the user to the network. Or, it could be the amount of traffic that the last point can see when sending traffic to its final destination. Thus, the difference in the design of the anonymity services regarding how to relay the traffic does affect the difficulty (anonymity) of correlating the user with their traffic.

4.2 Blocking Anonymity and Obfuscation Options

The anonymity systems can hide user activity on the Internet, but could not always hide the fact that such a system is in use. Sometimes, using anonymity systems might raise questions about why such a system is in use. In some countries, the IP addresses of the hosts running such systems are blocked, so obfuscation systems are used to evade these. Furthermore, Pluggable Transports (PT) [17] (obfuscation for Tor) work as an interface between the Tor user and the Tor network. The user connects to a pluggable transport, which sends the connection request to the Tor network on behalf of the user in order to hide the connection between the user and the Tor network. These tools work differently, using different techniques to resist different blocking methods. For example, Obfs3 [16] is one of the PT that obfuscate the Tor TLS to look like random strings, using another layer of encryption to wrap the TLS handshake used by Tor. Scramblesuit [19] is another PT designed to prevent active probing attacks.

JonDonym has two options for bypassing blocking of the service. The first one is using TCP/IP forward method, in which the user will use an encrypted connection to another user who has unblocked access to the JonDonym network. The second method is using Skype to tunnel the blockage of the JonDonym service, which is more reliable than using the TCP/IP forward method.

On the I2P network, there are no obfuscation options similar to Tor pluggable transports, and it is thus possible for an observer to collect the routers' IP addresses with a harvesting attack [6]. However, the I2P network implemented other improvements in the design of the transport layer to obscure the identification of the I2P network traffic. In addition, several obfuscation options are still considered by the developers of the I2P network, including using padding

techniques at the transport layer to achieve random length, studying the signature of the packet size distribution, and studying the technique used to block Tor.

In short, an observer, who wants to de-anonymize the user, needs to determine that the user is connecting to the anonymity service in the first place. Therefore, the existence of obfuscation options is a factor that should be taken into consideration to measure the level of anonymity.

4.3 Application and Anonymity

The common way to use an anonymity service is to use the default browser of the aforementioned services to browse the web. However, these anonymity services can be used with other applications in addition to web browsing. This requires the user to configure the application and the anonymity service to work together. The configuration for these applications is not that simple for non-technical users. When configuring any application to work with an anonymity service, it is important to fully understand how this application works to ensure that the user information is not leaked.

The configuration of the application and how the user sets the application on the anonymity network is an important issue. For example, the web browser contains many details other than what anonymity system the user is using. The anonymity tools aim to make their browsers undistinguishable to raise the anonymity level. Tor browser is a modified version of Firefox based on Mozilla's Extended Support Release (ESR) Firefox branch [15]. It includes HTTPS-Everywhere [7], NoScript, modifying some of the default Firefox settings, and the default extension settings. JonDoFox is the browser of JonDonym and is a modified version of Firefox [8].

Even when using the default browser for anonymity services, the correct configuration for the browser is important to ensure the safety of the user against many Internet websites that track their visitors. Moreover, some of the tools used by web sites could also identify the user or their behavior for the purpose of advertisements, collecting data for different types of studies, or building a database about the visitors of the website. Thus, the question to consider is: how such tools address the trade-off between browsing the websites with full offered services and preserving the anonymity of the users.

4.4 Authority and Logs

No doubt that the policy of the anonymity services about the cooperation with the authority (operator or regulator) and keeping logs affects the level of privacy. For example, JonDonym's agreement with the operators requires keeping no logs and not exchanging information between operators of the mixes. The reason behind this policy is that the identities of the operators are known, and they work according to the regulations in their own countries. Therefore, in JonDonym, there are several points that must be taken into consideration when evaluating the anonymity of such a system:

- The mixes that construct the path are fixed. That means knowing that the user employs one of these mixes, e.g., the last mix, implies knowing the first and the second mix.
- The number of mixes on the JonDonym network is very limited compared to the Tor network. On the JonDonym network, there are only nine cascades; six are operated by companies and three by individuals.
- The operators of these mixes are known and registered. They work according to the regulations of the authorities in their countries.

On the other hand, on the Tor network:

- The nodes that construct the user's path are not fixed. The user connects to three nodes that change periodically. Therefore, knowing that the user connects to a specific exit node does not necessarily imply knowing the first or the middle node.
- The number of Tor nodes is around 8000, which makes it relatively harder to get information about them.
- The operators of these nodes are not known. Tor does not require their users to identify themselves when offering to run a node. This might help to protect the operators' identities, but it does not guarantee that the operators are trusted.
- The nodes on the Tor network are supposed to be online as much as possible. However, there is no guarantee because most of these nodes are run by volunteers.

Furthermore, on the I2P network:

- The I2P user has the option to modify the number of routers used when exchanging messages. In addition, end-to-end encryption is used. The concept of garlic routing is also used when exchanging messages. This way, messages that pass through the routers are not distinctive, which means the purpose or the content of the messages cannot be extracted or inferred easily. For example, information such as whether the messages form an extension to the number of routers in the tunnel or if they contain data would not be extracted from the messages.
- The I2P network is decentralized, so there is no single point that is responsible for or represents the network.
- The user does not need to know all the routers in the network to be able to use that network's resources.
- I2P network's design is different from Tor and JonDonym; it is basically designed to provide a private network within the Internet. The number of outproxies is very limited. This also makes the browsing outside the network low compared to Tor and JonDonym. Therefore, the possibility that the user will frequently use the same exit point is high. This does not mean that it is a threat, but increases the probability of correlating the user with their traffic based on factors such as access time, duration, and the amount of data used.

Based on the above, what information the service provider (operator) has about the user and the operator's willingness to provide this information when asked to do so is also important in measuring the level of anonymity.

4.5 Threat Models

The anonymity services are built based on the separation between the user identity and the data sent or received by the user. One of the threats that such services face is somehow correlating the user data with the final destination data. Hypothetically, this may be possible by monitoring the first point in the anonymity network and the last point that connects the user with the final destination through data analysis. For example, the path on the JonDonym network is known, and if the attacker has the ability to monitor the traffic from the first mix and the last mix (out of the last mix), then connecting the users of this cascade and the amount of sent and received data may be possible. The path on the Tor network is not fixed, but the correlation is also a possible threat. To this end, there are studies on using marking techniques to trace user activities, but they are often limited to a specific user, a specific webserver, or even a specific exit node. The attacker could compromise both an entry node and an exit node, in which case the traffic out of the entry node is marked. The attacker then watches for the mark to appear at the exit node. On the other hand, the design of I2P network makes this kind of correlation a low threat. The path is not fixed or specified; users build inbound and outbound tunnels that do not count on the type of the router. All routers on the networks can be part of any path. The encryption mechanism provides for the confidentiality and the integrity of the messages. However, if the attacker has the resources to monitor all routers, then they may have enough data to discover paths.

As for the JonDonym network, this type of attack can target a mix server. A mix server has a limit on the number of users it can serve. The attacker could use this limit to break the anonymity of the mix server. If the attacker connects to a mix server to fill its capacity (n) to the point $(n-1)$ when the user connects to the only space left in the mix server, the attacker could then isolate and detect the user's traffic.

The threat models are not the same for all anonymity services; what is considered a threat to one service may not be applied to another anonymity service. Even when they share the same threat to a certain saturation point, the level of the risk is not always the same. Therefore, to measure the anonymity of any anonymity service, the threat model should be taken into consideration, based on the environment or the purpose for which the anonymity service is used.

In summary, evaluating the level of anonymity should be done in a comprehensive way that takes into consideration the purpose, the design, and the environment. Thus, these five factors: the level of information available to the service provider, the obfuscation options, application anonymity, the authority and the logs, and the threat models are used in this research to measure the effectiveness of any anonymity service.

5 Evaluation

This section discusses how the aforementioned factors can be used to measure the anonymity of Tor, JonDonym, and I2P.

5.1 Factor Calculation

As a first step, we quantify these factors, so they are grouped into three categories, Table 1. These categories (High, Mid, and Low) are converted into numerical values as 100, 67 and 33, respectively. These numbers are chosen to simplify the representation of the three categories into three intervals. The High range is between 68–100 and represented by 100. The Mid range is between 34–67 and represented by 67, and so on. The exception is for the obfuscation, where it is labeled as "Yes" or "No", depending on it is used or not. The reason is that some of the anonymity systems involve obfuscation techniques, and others do not. Therefore, the value is set to 100 (No) and 0 (Yes). The higher the values for these factors, the lower the anonymity level of the system. For example, a 100 in the Threat model factor is applied whenever the threat in the case under study is very strong (i.e., highly probable). The three categories are represented by 100, 67, and 33 as an approximation for High, Mid, and Low ranges. It is possible to expand this step to improve the accuracy of quantification of the factors by: (1) Instead of using three levels; the factors could be evaluated as a scale, for example, from 10 to 100, (2) Furthermore, each value on the scale could represent the level of the anonymity of the factor in a predefined way. This way, the value of the factors is determined more accurately. For example, if we apply the extended scale to the "Threat Models" factor, then the values will include more intervals from 10 to 100 instead of 33, 67, and 100. The threats or attacks on the anonymity systems should be ordered to match the scale from 10 to 100. This requires the study and evaluation of all possible threats on the anonymity systems and their applicability. This way, the scale has predefined values for every possible threat against the anonymity systems in the threat models factor. The same step could be used for the other factors.

Table 1. Proposed anonymity factors

Level of information	High	Mid	Low
	100	67	33
Obfuscation	Yes	No	
	0	100	
Authority and log	High	Mid	Low
	100	67	33
Application configuration	Low security configuration	Mid security configuration	High security configuration
	100	67	33
Threat model	Low cost	Mid cost	High cost
	100	67	33

5.2 Weight Calculation

Given that the weights of the factors may vary from one evaluation environment to another, quantifying these factors to measure anonymity is necessary but

insufficient by itself. Also, the weights of the factors have to be considered. Therefore, the "Pairwise Comparison" technique is employed to evaluate the weight of these factors. Each one of the factors is compared with all other factors; then, the weight for the factor is calculated based on these comparisons. The higher the weight of a factor, the more important it becomes for the anonymity of a given service. Calculating the weights is performed until all factors are evaluated comparatively, as shown in Table 2.

Table 2. Calculating the weights

γ_1				
γ_2	$2\ \gamma_1$			
γ_3	$\gamma_1\ \gamma_3$	γ_3		
γ_4	γ_4	γ_4	$\gamma_3\ \gamma_4$	
γ_5	γ_1	$\gamma_2\ \gamma_5$	$\gamma_3\ \gamma_5$	$\gamma_4\ \gamma_5$

The first column in the table represents the five factors. The first factor "level of information available to the service provider" is presented by γ_1. The second factor is presented by γ_2 and so on. The second column shows the importance of γ_1 compared to all the other factors. The third column shows the importance of γ_2 compared to all other factors except γ_1, this is because the comparison between γ_2 and γ_1 already done in the second column. The comparison continues till all the factors compared with each others. Each cell starting from the second column shows the result of the comparison between two factors; the most important factor is shown in the cell, however if both have similar importance then both appear in the cell.

Table 3 shows the weights of the five factors after the comparison and their total value. Based on the weights value, the level of information, the application configuration, and the authority and log factors have the same weights (importance). The obfuscation has the lowest importance, compared to the other factors. The weights represent the importance of each factor compared to the other factors. Using the pairwise comparison helps in deciding how to rank or weight the factors compared to others.

Table 3. Final weights of the factors after pairwise comparison

γ_1	γ_2	γ_3	γ_4	γ_5	Total
4	1	4	4	3	16

5.3 Weighted Anonymity Factor

Equations 1 and 3 are applied after calculating the values of the factors and weights. Equation 2 is the total of the weights of the factors. WF in Eq. 1 is the Weighted Anonymity Factor, (f) represents the value of a factor and γ represents

the weight. n in Eq. 2 represents the number of factors. T_γ in Eq. 3 is the total weight.

$$WF = \gamma_1 f_1 + \gamma_2 f_2 + \gamma_3 f_3 + \gamma_4 f_4 + \gamma_5 f_5 \tag{1}$$

$$= \sum_{i=1}^{n} \gamma_i \, f_i \tag{2}$$

$$(T_\gamma) = \sum_{i=1}^{n} \gamma_i \tag{3}$$

The measurements may vary from one environment to another, where different factors are applied or when the numerical conversion is different than what is used in Table 1. To generalize measurements, Eq. 4 shows converting the calculated values of the weighted factors (WF) from Eq. 1 to a percentage by using the minimum and maximum value.

$$WF\,(\%) \; = \; (1 - \frac{WF - Min(WF)}{Max(WF) - Min(WF)}) * 100 \tag{4}$$

Equation 4 can be rewritten after calculating the weights to the form in Eq. 5.

$$WF\,(\%) \; = \; (1 - \frac{WF - 495}{1600 - 495}) * 100 \tag{5}$$

5.4 Evaluation Case Study

In this case study, three users participate to compare the levels of anonymity. It is important to note that the evaluation does not aim to identify the best anonymity service; it aims to evaluate the level of anonymity according to the environment in which these users use the anonymity services.

The first user (A) uses standalone Tor to browse Internet websites. She configures Chrome browser to work with Tor by setting the browser to access Tor via Socket. To increase the anonymity level, she adds Scramblesuit as an obfuscation option to her "torrc" file to access Tor via a bridge. She browses websites on the Internet which include a compromised web server by an attacker. The web server injects a code to force the browser to request images from another website that belongs to the attacker. The attacker aims to identify the user by forcing the browser to send requests without using the Tor network.

User (B) chooses to use JonDonym as an anonymity service. He does not have a technical background. All the settings are left as default. The only addition to the default setting is that he chooses to use the TCP/IP forwarder. The user (B) wants all the activities that he performs on the Internet to be anonymous. Therefore, he uses JonDoFox to browse all the Internet websites. He usually visits web sites such as news, videos, email, Internet shopping, and his bank account.

User (C) lives in a country where the Internet is censored and some websites are blocked. Therefore, he uses Tor to gain access to the blocked Internet blogs. The user (C) browses these blogs and participates on them via Tor. He is concerned about hiding his identity, so he uses the Internet via the company

network where he works. It seems that he is the only person who is using Tor in this company. The user organizes his time so that he only accesses Tor at the end of the day between 5–6 pm on weekdays.

Table 4. Evaluated factors for users (A), (B) and (C)

	Level of information	Obfuscation	Authority and log	Application configuration	Threat model
A	33	0	33	100	67
B	100	0	67	33	100
C	67	100	100	33	67

Table 4 shows how these scenarios are converted to measurable numeric values, using the proposed factors. Table 4 is calculated based on the given information about the scenarios above and how the users (A), (B) and (C) are using these anonymity services. For example, the user (C) did not include an obfuscation option when using the anonymity service; therefore, the obfuscation value is measured as 100. The user (A) prefers to use his favorite browsers instead of using the default Tor browser. Therefore, the possibility of a DNS leak is higher, especially when accessing suspicious websites or when using any application other than browsing. Based on that, user (A) gets 100 on the application configuration. Even though the user (B) uses some sort of obfuscation, he misses the fact that browsing any website already linked to his real identity such as his email or bank account, even while using an anonymity service, does not mean that he is anonymous. Furthermore, the information available to the exit node in this case is high, even if the information does not contain passwords. So, the level of information is evaluated as 100 in this case. The same applies to the user (C); he uses Tor at the same time daily from the same place where no one else is using Tor. Using Table 4 and Eq. 1, the weighted factors are calculated as follows:

$$WF = \gamma_1 f_1 + \gamma_2 f_2 + \gamma_3 f_3 + \gamma_4 f_4 + \gamma_5 f_5$$

$$WF = 4f_1 + f_2 + 4f_3 + 4f_4 + 3f_5$$

$$\begin{aligned} WF_A &= 4 * 33 + 0 + 4 * 33 + 4 * 100 + 3 * 67 \\ &= 865 \end{aligned}$$

$$\begin{aligned} WF_A(\%) &= (1 - \frac{865 - 495}{1600 - 495}) \\ &= 66.5\% \end{aligned}$$

$$\begin{aligned} WF_B &= 4 * 100 + 0 + 4 * 67 + 4 * 33 + 3 * 100 \\ &= 1100 \end{aligned}$$

$$\begin{aligned} WF_B(\%) &= (1 - \frac{1100 - 495}{1600 - 495}) \\ &= 45.2\% \end{aligned}$$

$$\begin{aligned} WF_C &= 4 * 67 + 100 + 4 * 100 + 4 * 33 + 3 * 67 \\ &= 1101 \\ WF_C(\%) &= (1 - \frac{1101 - 495}{1600 - 495}) \\ &= 45.16\% \end{aligned}$$

Based on the above calculations, user (A) has a higher level of anonymity than either of users (B) or (C). The level of anonymity of user (A), (B) and (C) may change based on the anonymity services they use or even their behavior. The weighted factor method is designed to take into consideration the users' environment when evaluating the anonymity level. The factors themselves are parameters that could be adjusted based on the scenario for which an anonymity system is used.

6 Conclusion

In this paper, we propose and evaluate five factors that affect the level of privacy in anonymity services. Understanding these factors and knowing how to address them is an important step in improving users' privacy. To this end, three popular anonymity systems, namely Tor, JonDonym, and I2P, were used as case studies to analyze these factors. Our analysis showed that even though these systems aim to provide anonymity to their users, user information is visible to the operators of the services. Furthermore, the infrastructure and the browser settings vary from one system to another. The setting is configured based on the developers'/administrators' evaluation of possible threats. The same threat might be considered high in one system but low in another. We applied a measurable mechanism to evaluate the anonymity of a given situation based on the factors we proposed. This evaluation could be used on any anonymity system using different scenarios. Future research will continue to analyze other anonymity systems based on the proposed five factors and will evaluate them using the proposed approach under other adversarial conditions.

Acknowledgment. This research is partially supported by the Natural Science and Engineering Research Council of Canada (NSERC) grant, and is conducted as part of the Dalhousie NIMS Lab at http://projects.cs.dal.ca/projectx/. The first author would like to thank the Ministry of Higher Education in Saudi Arabia for his scholarship.

References

1. Dhiah el Diehn, A.-T., Pimenidis, L., Schomburg, J., Westermann, B.: Usability inspection of anonymity networks. In: 2009 World Congress on Privacy, Security, Trust and the Management of e-Business, CONGRESS 2009, pp. 100–109. IEEE (2009)
2. Chaum, D.: The dining cryptographers problem: unconditional sender and recipient untraceability. J. cryptol. **1**(1), 65–75 (1988)

3. Clark, J., Van Oorschot, P.C., Adams, C.: Usability of anonymous web browsing: an examination of Tor interfaces and deployability. In: Proceedings of the 3rd Symposium on Usable Privacy and Security, pp. 41–51. ACM (2007)
4. Díaz, C., Seys, S., Claessens, J., Preneel, B.: Towards measuring anonymity. In: Dingledine, R., Syverson, P. (eds.) PET 2002. LNCS, vol. 2482, pp. 54–68. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36467-6_5
5. Garlic routing (2014). https://geti2p.net/en/docs/how/garlic-routing
6. I2P's threat model: Harvesting attacks (2010). https://geti2p.net/en/docs/how/threat-model
7. HTTPS-everywhere extension. https://www.eff.org/https-everywhere
8. Anonymous surfing with JonDoFox (n.d.). https://anonymous-proxy-servers.net/en/jondofox.html
9. Jondonym InfoService (n.d.). https://anonymous-proxy-servers.net/en/help/infoservice.html
10. Maymounkov, P., Mazières, D.: Kademlia: a peer-to-peer information system based on the XOR metric. In: Druschel, P., Kaashoek, F., Rowstron, A. (eds.) IPTPS 2002. LNCS, vol. 2429, pp. 53–65. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45748-8_5
11. McCoy, D., Bauer, K., Grunwald, D., Kohno, T., Sicker, D.: Shining light in dark places: understanding the Tor network. In: Borisov, N., Goldberg, I. (eds.) PETS 2008. LNCS, vol. 5134, pp. 63–76. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70630-4_5
12. Murdoch, S.J.: Quantifying and measuring anonymity. In: Garcia-Alfaro, J., Lioudakis, G., Cuppens-Boulahia, N., Foley, S., Fitzgerald, W.M. (eds.) DPM/SETOP -2013. LNCS, vol. 8247, pp. 3–13. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-54568-9_1
13. Ries, T., Panchenko, A., Engel, T., et al.: Comparison of low-latency anonymous communication systems: practical usage and performance. In: Proceedings of the Ninth Australasian Information Security Conference, vol. 116, pp. 77–86. Australian Computer Society, Inc. (2011)
14. Serjantov, A., Danezis, G.: Towards an information theoretic metric for anonymity. In: Dingledine, R., Syverson, P. (eds.) PET 2002. LNCS, vol. 2482, pp. 41–53. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36467-6_4
15. The design and implementation of the Tor browser (2017). https://www.torproject.org/projects/torbrowser/design/#idm29
16. Tor Obfs3 (n.d.). https://gitweb.torproject.org/pluggable-transports/obfsproxy.git/tree/doc/obfs3/obfs3-protocol-spec.txt
17. Tor pluggable transports (n.d.). https://www.torproject.org/docs/pluggable-transports.html.en
18. Wendolsky, R., Herrmann, D., Federrath, H.: Performance comparison of low-latency anonymisation services from a user perspective. In: Borisov, N., Golle, P. (eds.) PET 2007. LNCS, vol. 4776, pp. 233–253. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75551-7_15
19. Winter, P., Pulls, T., Fuss, J.: ScrambleSuit: a polymorphic network protocol to circumvent censorship. In: Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society, pp. 213–224. ACM (2013)