Alex Shah
EN.695.741.81.SP25 Information Assurance Analysis
Mod 5 Assignment
February 24, 2025

1. There are tradeoffs against anonymity and identifiability, discuss what some of those are?

Being able to find your own information on websites like Spokeo is disconcerting. With so many websites and services, you can play whack a mole trying to find and remove your personal information. Removing your information and trying to remain anonymous could help prevent harassment online and in the real world. However there are some benefits to making people accountable online by identifying them. For example, totally anonymous activity can make people feel secure behind a mask and attempt to harm others. There are clear reasons to want privacy by being anonymous but also reasons against people being able to hide their identity. State actors, companies, and individuals are capable of tracking or finding personal information online which makes the tradeoffs multifaceted.

Anonymity provides some protection against surveillance and the ability for free expression online, including more serious topics like whistleblowing and journalism in areas that limit speech. It also allows for access to censored information, for example getting around China's firewall with an anonymity tool like Tor. However this level of anonymity can also enable bad actors like hackers and spreaders of misinformation who hide behind anonymity. Good anonymity makes it difficult for law enforcement to conduct investigations and for cybersecurity and IT administration to prevent bad actors from accessing sites and infrastructure. Reasonable privacy and law enforcement efforts are both valid concerns and both need to be balanced for everyone's safety on and offline.

There are also financial and security interests in big data and the ability to keep track of user's movements on the web for behavioral insights or just selling advertisements. With so many companies' revenues built on the collection of specific users' data, it is unlikely that large companies would allow their collection of data to be hampered by anonymization efforts more than they have to. The same is true of governments that also want to leverage large scale data collection to analyze patterns and find more fine grained insights. Anonymization tools obscure the inferences that governments might rely on for security.

2. Do you believe that anonymity can be achieved technically, or is there a legislative/political component to anonymity?

Anonymity is becoming increasingly complex technically and by legal and government intervention in recent years. Legislative and political forces have been limiting the effectiveness of technologies and tools that enable anonymization. Governments around the world want to break encryption, introduce backdoors, and create mandates for gathering and storing user data that impact the ability to remain anonymous online.

There will probably continue to be tools that provide a layer of anonymization for online activity. But these tools could become harder to access or be less useful if there is increased focus around breaking anonymity. For example, requiring identification to access certain resources, or the use of backdoors, fingerprinting, and stringent data collection. I would imagine means like PGP and bespoke end to end encrypted applications could still allow anonymous message passing, but it's likely that with recent developments in some countries that legal pressure to use backdoors could compromise

end to end encrypted apps and infrastructure. For example, recently in the UK Apple was compelled to install a backdoor into encrypted iCloud operations, and they responded by removing end to end encryption access in the UK. Even without direct backdoors or requiring identification, methods like browser fingerprinting, embedded trackers, and more in depth traffic and behavior pattern analysis could be used to break anonymity despite use of privacy tools. Especially with the resources and volumes of data collection that a government might have.

I would anticipate that the legal fight against concepts like encryption will be ongoing, which have a crucial role in protecting anonymity and privacy online. However there are also good privacy laws like the GDPR that protect user privacy by limiting data collection and describing how it can be used, though measures like these are not widespread or the norm.

3. Should anonymity be required in all network interactions, or in specialized ones only?

Anonymity would not benefit certain network interactions. Requiring it would either not be beneficial or in some cases it may be detrimental or impractical. There are lots of reasons to need to identify specific users like banking and social media. If anonymity were required for banking interactions, it would be difficult for users to prove their identity to access their accounts. Or it may make it easier for someone to masquerade as someone else. Anonymity on social media and other contexts would also provide a breeding ground for misinformation and bad actors to flourish. In some cases anonymity could make it more difficult for legal investigations such as looking into online criminal activity and national security interests.

However anonymity is still necessary in some applications to protect users' privacy. For example it is important for anonymity to remain viable for journalists and whistleblowers to share sensitive information without fear. In authoritarian or censoring governments, anonymity enables its users to dissent or protest and otherwise freely express themselves online. So there are special interactions where anonymity is crucial, but using it for all network interactions presents some problems. There could be compromises, like allowing anonymity in the right contexts like some social media or communication apps. But anonymity should not be required for all network interactions.

4. The Onion Router (ToR) is designed to provide anonymity rather than confidentiality. How does ToR establish anonymity for the user? (reference the last three attached readings)

Tor provides anonymity rather than confidentiality in dealings online by routing traffic to obscure the users' identities and locations. Tor uses multiple hops and an onion routing scheme to encrypt packets with a shells of encrypted layers which prevents any node from having access to the content's of the message until it reaches its destination. While it does protect the message with encryption layers, the scheme is designed to obfuscate traffic and prevent any point in the network from knowing the destination and source of the packet. The routing path is also changed regularly to prevent analysis of traffic patterns to determine the origin or path of particular packets or connections. This decentralized method makes users' traffic anonymous and resistant to censorship and surveillance, provided that the user protects their private information while using it, such as not logging into a service which would identify them. Otherwise Tor makes it very challenging for an adversary to trace the user's activity and connections.

In onion routing, when a user sends data through Tor it is encrypted multiple times like layers of an onion. The data passes through multiple nodes on the network where one layer is decrypted and the next destination hop is read. In this way, the node only sees where the packet came from on its last hop,

and where it is going next. This ensures that no node can see both the source and destination of the real connection. At a minimum this usually involves 3 nodes, an entry point into the network, a relay in between, and the exit node that connects to the intended endpoint. The entry node may know the real source IP address, and the exit node may know the real destination IP address, but no node knows them both. The node in the middle only knows the hops before and after it as source and destination. Even if an adversary were to be in control of a node, they wouldn't be able to piece together enough information to find the true identity or destination of the user's activity.

There are also similar mechanisms like VPNs that route traffic through a middle man so the true identity of the source doesn't reach the destination. However, the centralized VPN servers and single point of control make it easier to compromise anonymity, especially in instances where the VPN is required by law to turn over records or an adversary is able to surveil the VPN. There is only one hop between the user and the destination, meaning watching a single node can reveal information about the real source and destination, Tor on the other hand is a distributed network hosted by volunteers that uses multiple hops. This makes it harder for governments or adversaries to watch traffic or shut down connections. Tor also enables a feature to disguise traffic between the user and the entry node by using bridges that can look like regular traffic. Some of these bridges are unpublished so that censoring governments and other entities cannot block all the bridges they are unaware of. In areas with heavy restrictions, someone might use a bridge to disguise or randomize their traffic as they enter the Tor network, then proceed through multiple hops and access censored data. This protects their anonymity in areas where Tor itself is restricted.

5. From the readings, "The Challenges of Effectively Anonymizing Network Data", Section 3.1 Anonymization Methods, talks about four techniques to achieve this; truncation, randomization, quantization, and pseudonymization. Pick three (3) of these, do some additional research and explain how the method works?

Truncation anonymizes data by removing portions of identifying information like part of an IP address in order to provide some information but not enough specificity to identify an individual. Another example might be an approximate age range instead of including someone's birthdate. This is a good middle ground for preserving some information while limiting identification risk. It is also an easy system to implement but it still requires some careful thought, since there might still be a risk for identification given enough approximate factors. This does come at the tradeoff to accuracy when using the data such using ranges, which may not be sufficient for stringent research.

Randomization introduces noise or modifies fields with randomized data in order to obscure it. For example, timestamps or GPS values can be offset by random intervals to prevent patterns that reveal someone's habits. These random offsets change the data so that it can't be used to identify someone, in a way that is less detectable to an adversary and harder to account for. This could also cause data to be less accurate, such as randomized GPS data providing locations of limited trust value. This helps reduce the confidence an adversary has when accessing a dataset. The random and unpredictable aspect to the data offset helps make it irreversible to attackers and obscures patterns but like truncating, it reduces the accuracy of the data and its usefulness. Adding random noise to data may not alter the underlying statistical patterns of the data which may make some patterns and predictions still feasible, however.

Pseudonymization can provide anonymization by replacing real identifiers with made up data or fake names. For example, a random identifier can be used instead of someone's identifying name or social security number in a dataset. Pseudonymization is often not truly anonymous since it is possible

to link pseudonyms with real information and use that to discover real identities. Some compliance methods like GDPR use pseudonymization to maintain the possibility of reidentification. And pseudonyms can still be used internally to track users for example at a retailer. Pseudonymization helps protect real identifiers while also allowing access to real data that may be more useful for analysis, at the cost of potential reidentification which makes it not truly anonymous.

6. Is anonymity an achievable event or a futile pursuit due to advances in computing, increased attack surfaces/vectors, third party proxies, network port scanning, passive/active collection methods, and the use of digital data making anonymity difficult across platforms and applications?

Like most cat and mouse games, anonymity is made more difficult by advances in technologies, but also evolves along with them. It is especially difficult in the age where data and tracking are important to both governments and companies. While laws and technical forces may make anonymity more difficult to achieve, it is not futile. It should still be possible in the future with the right tools, effort, and vigilance to maintain anonymity online.

Some of the biggest hurdles to anonymity are the analytics surrounding data gathered while trying to remain anonymous. Machine learning techniques and large datasets have led to strong pattern recognition especially when coupled with new techniques for tracking movements and behaviors online like mouse movements and typing patterns. Some of the most innocuous details reported by your browser can be used for fingerprinting a specific user, such as the installed fonts on your browser.

There are also many more devices coming online such as through the Internet of Things and more connections between systems that make fingerprinting and identification much more likely. Through the complex interactions of systems and programs, there is an increased attack surface and limited ability to disconnect from the devices and services that could identify you. Common techniques for anonymization have also been compromised or have vulnerabilities that well funded actors have taken notice of to thwart anonymization. For example Tor exit nodes in control of an adversary may reveal information about the plaintext interaction with the destination, which could reveal information about the user. There are also traffic correlation attacks where watching the entry and exit nodes in Tor can reduce how effective they are at anonymizing traffic. Governments and new laws have been scrutinizing the use of anonymization and encryption methods as criminal tools. Some countries are pushing for backdoors, more data collection, and surveillance compliance such as by ISPs and social media platforms. Some countries even ban anonymity tools like Tor is banned in China.

However, anonymity is still possible. The bridges mentioned earlier can be hidden or new ones can pop up to allow users in places where Tor is banned to still access the protocol. There are also lesser used peer to peer networks or protocols that are less scrutinized or not explicitly banned that users can switch to. There are also new crytpographic methods like homomorphic encryption that allows for transformations on encrypted data and differential privacy. This could allow anonymization to be preserved while also protecting the online and real world interests that have been scrutinized with strong end to end encryption. There are also laws that specify how data can be collected from users like the GDPR and CCPA laws in the EU and California that reign in data collection methods. These are privacy preserving laws and initiatives that show that even in the legal system, the future isn’t entirely bleak for anonymization and overreach can be curtailed with privacy laws.

7. What are other technologies/methods for implementing anonymity on a network? (i.e., mix networks, batching strategies, high-latency, low-latency anonymity systems, etc.)

There are multiple methods for implementing anonymity on a network that come down to technologies and strategies. Mix networks introduce delays and reorder transmissions while sending traffic through multiple nodes in order to make it more difficult to perform timing analysis on traffic to detect patterns. Mix nets can also vary delay and batch messages together in order to further confuse timing analysis. Higher latency systems like mix nets can provide higher levels of anonymization by sending traffic on longer or more confusing routes which increases time delays for the user. For more critical uses of anonymization, the trade off can be worth the inconvenience. For example in email based applications with a high latency system, messages can be exchanged with some delay while enhancing privacy and anonymization efforts by bouncing the messages around. However for real time applications or even the responsiveness of interacting with webpages, high latency mix nets are not ideal. The network also requires trusting the nodes that route traffic, and like Tor there are vulnerabilities when the nodes are compromised or many are being watched.

Low latency systems like Tor prioritize speed over anonymity by sending traffic through multiple hops without introducing long delays or waiting to batch messages. They can provide users with a reasonable amount of anonymity for low stakes applications by routing through multiple nodes on the path to the destination and using encryption like onion routing to enhance anonymization efforts. Tor is better suited to web browsing and closer to real time interactions because there aren't any added delays other than the time to route through the nodes. This makes low latency systems more susceptible to timing analysis, especially if nodes are compromised or entry and exit nodes are being surveilled.

There are also newer or more cutting edge methods to retain anonymity without revealing information. Homomorphic encryption allows for operations on encrypted data without the need for decryption. However it does take more space and computational resources than if the messages were decrypted first. There are also zero knowledge proofs that allow an entity to prove they know something without revealing what the information actually is. An entity in a ZKP system might be challenged to use a value they know to generate a response that satisfies a verifier, for example using a value to compute a hash, where the prover would only be able to do so if they actually knew the value. In this way the verifier can be sure that the prover knows the value, without ever exchanging the value or sending any sensitive information in an encrypted form. This system can provide privacy preserving verification without the need to give away data.

8. How does Zero Trust Architecture impact the ability to implement and maintain anonymity (reference the attached NSA Security Model pdf- found in this week's Readings)

Zero trust architecture assumes that no user or device can be trusted, and enforces constant verification, access controls, and monitors all interactions as a security model such as in a critical system where personal or medical data might be used. It enhances security, but makes it more difficult to remain anonymous. With zero trust models, proof of identity is a strict requirement. Every device needs to prove their identity before gaining access to a resource, which might even use multi factor authentication or biometrics to verify. This would directly give away the identity of the device and user. Trying to avoid identifying oneself would prevent access to resources since the strict access control policies would prevent unauthorized devices, which makes anonymity impractical in a ZTA.

ZTA also follows the principle of least privilege which prevents entities from seeing anything they are not explicitly granted permission to access. The system is designed to give only the minimum necessary access to data for a device to perform its function. And with the strict tying together of identity and access, refraining from identifying oneself to remain anonymous would mean that access

to resources would be minimal or nonexistent. Since ZTA segments resources based on identity, there would be no way to traverse or access data.

ZTA also relies heavily on logs and monitoring that would create trails when someone wanted to access or move through a system anonymously, which further hampers anonymization efforts. Data collection about the interactions in a system could compromise the anonymity of a user of a ZTA system by including details like timestamps, resources accessed, and authentication events necessary to access resources. Aside from directly proving identity to use the system, behavioral or statistical analysis can identify users by the data they attempt to access and other factors logged along with the interactions with the systems. The nature of ZTA systems also attempts to coordinate dynamic and static information about every device and event to form context to allow or deny access, shining a spotlight on all the characteristics and movements a device attempting to remain anonymous might create in a log when accessing a ZTA system. These logs are used to inform decisions before every action. So ZTA is inherently based on principles that make anonymity difficult if not impossible to achieve in order to access resources on the system or move within its perimeter.

Sources

Dingledine, R., Mathewson, N., & Syverson, P. *Tor: The second-generation onion router.* USENIX. https://www.usenix.org/conference/13th-usenix-security-symposium/tor-second-generation-onion-router

Farah, T. (n.d.). ALGORITHMS AND TOOLS FOR ANONYMIZATION OF THE INTERNET TRAFFIC. https://www.sfu.ca/~ljilja/cnl/pdf/Thesis_tanjila_final.pdf

Narayanan, A. (n.d.). Robust de-anonymization of large sparse datasets. https://www.researchgate.net/publication/4339941_Robust_De-anonymization_of_Large_Sparse_Datasets

NSA issues guidance on Zero trust security model. National Security Agency/Central Security Service. (2021, February 25). https://www.nsa.gov/Press-Room/News-Highlights/Article/Article/2515176/nsa-issues-guidance-on-zero-trust-security-model/

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020, August 11). *Zero trust architecture*. CSRC. https://csrc.nist.gov/pubs/sp/800/207/final

https://www.ibm.com/think/topics/homomorphic-encryption

https://umatechnology.org/what-is-a-zero-knowledge-proof-in-cybersecurity-and-how-does-it-work/

https://support.torproject.org/about/protections/