21st Cyber (Necessary Algorithms): Consensus Protocols

LIS461: Data and Algorithms: Ethics and Policy

Shia Aaron Lloyd Fisher

University of Wisconsin, Madison

18 July 2023

Shia Aaron Lloyd Fisher

Paul J. Kelly, Tallal Ahmad, Yinka Ajiobola

Data and Algorithms: Ethics and Policy – Paper #1

18 July 2023

Consensus Protocols

It is necessary for an algorithm that implements a consensus protocol to ensure agreement within the systems' objects in memory. A consensus protocol not only ensures agreement, but it also can measure the level of validity for a particular object or object's proposed value in data. Mainstream systems used for financial records manage hundreds of thousands of transactions daily. These transactions are large and small and can affect the lives of many people. Hence it is required for financial software to implement algorithms that are secure, accurate and reliable. This can be achieved using the Borowsky-Gafni (BG) algorithm.

The reason it is important for financial software to have secure, accurate and reliable data transmission is obvious. Without a secure system customers would risk losing their real financial property in the event of breach. According to Statista, there has been a steady increase since 2005 in the number of annual data compromises in the United States with 1,802 data compromises occurring in 2022 affecting over 422 million individuals. This includes events outside of the financial sector but always involves sensitive data, nevertheless.

Another reason for financial software to have accurate and reliable data transmission is that it is matter-of-factually the purpose of said software. If a financial software developer were by counter-example unable to guarantee accuracy or reliability, that developer would go out of business since no bank would buy their software. Therefore, the worry for the developer rests solely on using algorithms that protect the individual customer's personal information while relying on information to verify said individual's credential, satisfying the security requirement.

Not all financial software uses BG, it may also be the case a financial software is able to ensure security, reliability, and accuracy without the use of a consensus protocol altogether. However, consensus protocols are ostensibly required for some of the leading distributed systems used by millions of people daily, for instance cloud-computing, blockchain and other multi-agent systems. In September of 2020, software developer Oracle, employer of 164,000 employees, implemented consensus protocol inside their Blockchain product (Macrotrends 2023). Oracle Blockchain Platform's Principal Architect Baohua Yang credits the reason for consensus protocol's popularity is due to the level of significance it offers a system if developed and implemented correctly, otherwise known as scalability. In his 2020 blog post he offers some heuristic for deciding which consensus protocol to implement. Consensus protocols should be implemented in all distributed systems that transmit or receive sensitive information. This paper will discuss the ethical considerations of implementing the BG consensus protocol.

One key concept to grapple with is distributed systems. Distributed systems have an evolving history spanning over 60 years and covering the range of Internet of Things (IoT), to gaming, networking, world wide web, etc. (Lindsay 2021). Diving further into the concepts used in consensus protocols, are pieces of data that might constitute, for instance, a user's account information. This may include dates of birth, banking information, residential addresses, or other demographic information. To adopt a deontological framework for this software development, one must consider the autonomy of the audience that comes into contact with a particular software. If the software is online, it may be possible for anyone to come into contact with the software.

In 2013 the state of California imposed the California Online Privacy Protection Act (CalOPPA), first proposed in 2003 (Harris 2013). This law requires businesses to provide public notice on their online sites to articulate the data collection, use, and retention policy for their site. Even a user who does not utilize a particular software should have their rights respected in terms of personal online activity as this is approximately the limit of exposure one assumes by accessing a website.

For users who intend to use a service, they may be exposed to greater data collection. Many websites to register require a birthdate, and a valid email at the minimum to obtain a user account. While this information may be used for some other purpose, this information, when offered together with other personal information, may be used to validate, or credential a user to access additional privileged information or place transactions. In the case of managing an online bank account one’s birthdate, address, social security number, telephone number are all ways a user’s account can be identified. California law provides a certain level of informed consent to the public, further use of a particular software will (likely) result in the collection of additional information and informed consent. Whereas informed consent is a cornerstone of the Kantian principle that respects autonomy, California law protects online users from sharing data they do not wish to share.

Another key concept here is the multiplicative inverse. This concept is widely discussed in cryptography which is a topic of discrete mathematics. One effortless way to understand is to consider, for example, a set of ordered prime numbers that are individually assigned to users in a particular financial system. This number, in conjunction with a verifiable username and matching password credential a user's login attempt, however the number itself is not made known to the user nor to any user of the system. The algorithm checks the value of the user's account which is an object classified by the system. An algorithm that takes part of the input data stream, say the username and password, and applies a multiplicative algorithm to both the input and output will result in the greatest common factor of one ("1") if the input and output are the same. Thus, allowing the user to gain additional access to that account.

Let us assume there exists a particular financial system which runs its own "distributed system." Data about its individual users are stored in hard drives. By tracking the number of processes, value of data, and number of errors, BG could use the mathematical concept known as the multiplicative inverse to verify if the value assigned to one data stream matches another one. That is, two objects are identified as one in the same if and only if their consensus numbers match. With the BG algorithm, only consensus numbers need to match, which means other information that may be private is hidden. Since a consensus number is a mathematically arbitrary number, it is not possible to re-identify a particular user within a system by their consensus number alone. This implies security for a multi-agent system since information fed in cannot be back tracked or reversed to gain access.

This clever deployment of a mathematical principle is wise since the developer can tailor the algorithm to use the user's most sensitive input data. In the case of banking, financial institutions may want to guarantee only one user account exists for one individual. Requesting a social security number upon registration may allow the system to verify if a matching social security number was previously registered. However, it may also be the case a user could have multiple accounts, so other differentiating inputs may be useful. Since social security numbers are unique (meaning they are not repeated or shared by other US citizens), it can be used to encrypt a data stream. The number can be mutated to be a multiplicative inverse of an arbitrary user ID to generate a unique prime number inside a distributed system. Consequently, affording a user the ability to recover their account without necessarily revealing the true user ID number. The object containing the user's account information instead only contains the encrypted version of the user's ID which is useless without the (d)encryption algorithm.

We have already broached the ethical concept of "moral autonomy," which is the state or condition of self-governance. We showed that a website that makes every reasonable attempt to inform the user of data collection and safeguards privileged information from public view grants a certain level of autonomy. One other relevant ethical concept to consider is "moral legitimacy," which begs to question whether or not a software is abiding of laws, and for the sake of argument, supposes an algorithm is free from errors that may result in data leakages or breaches. To make progress toward developing a financial software that incorporates consensus numbers and is legitimate, a baseline set of standards must be articulated and accepted by the development community and/or law makers.

Law makers have the ability to require corporations to change business practices as demonstrated by CalOPPA. Indeed, many financial institutions set out with unambiguous language, the ethical standard for their agents, prohibiting even the appearance of impropriety. It was previously mentioned the level of negative impact caused by individually compromised accounts. Therefore, it is not only morally acceptable to require better encryption standards such as BG, but it may also be morally required for lawmakers to modernize legislation to do just that.

Already the government requires adherence to the National Institute of Science and Technology (NIST) standards and guidelines before even considering a software developer's bid for contract work. BG is an abstract, reduction algorithm that can be understood by software engineers and academics. There exist several completed algorithms that use consensus numbers to ensure legitimacy in data transmission. These other consensus algorithms may act similar to the BG abstract. One key characteristic of the BG algorithm that should specifically intrigue law makers is that it counts the number of faults per object. This means the system is able to canonically log the number of failed access attempts. This fault too can be used by the algorithm, thereby mutating a user's hidden account number more often. This too acts as a safety feature, saving a compromised system from remote attacks.

Finally, it is not expected the typical user recognizes the many facets related to their privileged information being stored in various distributed systems. Afterall, one does not need a law degree or computer science degree to check their bank statement online. There is a simple set of expectations these users have (legitimacy, security, etc.). It is incumbent on lawmakers to impose and enforce regulations that ensure constituents operating the public spaces or marketplaces are safe to make digital transactions without compromising privileged information or risking loss.

References

Bitcoin Transactions Per Day. Ycharts. Retrieved July 18, 2023, from https://ycharts.com/indicators/bitcoin_transactions_per_day

Imbs, D., & Raynal, M. (2010). The multiplicative power of consensus numbers. Proceedings of the 29th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing, 26–35. https://doi.org/10.1145/1835698.1835705

Lindsay, D., Gill, S. S., Smirnova, D., & Garraghan, P. (2021). The evolution of distributed computing systems: From fundamental to new frontiers. Computing, 103(8), 1859–1878. https://doi.org/10.1007/s00607-020-00900-y

Harris, K. D. (2014, May). Making your privacy practices public - state of California. https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/making_your_privacy_practices_public.pdf. Accessed 18 July 2023.

Oracle: Number of employees 2010-2023: ORCL. Macrotrends. (n.d.). Retrieved July 18, 2023, from https://www.macrotrends.net/stocks/charts/ORCL/oracle/number-of-employees.

Petrosyan, A. (2023, April 1). Number of data breaches and victims U.S. 2022. Statista. https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/

Yang, B. (2020, September 29). Oracle Blockchain Platform adds the Raft Consensus Algorithm. Oracle Blogs. https://blogs.oracle.com/blockchain/post/oracle-blockchain-platform-adds-the-raft-consensus-algorithm