INTEGRA

Integra Ledger proposed Blockchain Infrastructure

David Berger

05/24/18

The legal industry is plagued with a staggering data reconciliation issue with each law firm and legal department utilizing its own isolated systems to track common data and IDs from legal matters and contracts to lawyers and law firms. The existing solutions also tend to lack in transparency. Given the often $600 + an hour billing rate, this has led to issues ranging from large-scale inefficiencies to worldwide access to justice concerns. This stems from a combination of social and technological factors. There are many existing potential solutions utilizing existing technologies but most are at least partially problematic. We propose that Blockchain and related technologies are an excellent candidate for closing the remaining gaps.

However, it is important to be cautious as Blockchain is still in its infancy and is suffering from rapid change and lack of standards contributing the kind of chaos that is anathema to the legal mindset. As a result, the first adopters could potentially shoulder a disproportionate share of the initial cost. We refer to the resulting lack of innovation and paralysis as the “Collective Action Problem” which manifests in a fear of “Vendor Lock in”.

In an attempt to mitigate these issues, it is critical that any new proposed architecture changes be implemented in an incremental fashion in which the changes coexist with the existing systems. This reduces risk as it is easier to test new functionality in isolation allowing for immediate benefit which leverages network effects as adoption grows.

The exponential growth of legal data along with the complexities of emerging technologies such as AI has spawned a movement where the well-established practices of software engineering are being applied to the legal context beyond of the application of traditional legal tech. We have decided to focus on selected lessons of the widely accepted architectural patterns such as Service Oriented Architecture (SOA) and the related Micro Service architecture which has been successfully battle tested by Internet giants such as Amazon and Netflix.

In the spirit of incremental improvement, we have decided to focus on tacking a triangle of foundational issues: legal artifact identity, proof of existence of legal entities (hashes), and document/communications security. These issues are apparent in many common real-world scenarios.

Within a legal context, activity tracking is much more efficient if every aspect is tied to an anchoring entity, such as a matter with a unique matter ID, and each related artifact is also assigned its own ID. There are many enterprise grade matter management systems, too many in fact. Unfortunately, this lack of a common standard plunged the legal industry into a rat’s nest of independent silos with each law firm and legal entity, including the courts, running their own independent systems. Therefore, a single matter or artifact, such as a contract, could have dozens of different identifiers. Inefficiencies spawning from data reconciliation issues drives the cost of law significantly higher than it should be and is a prime example of the first point of the triangle we will address: legal artifact identity.

Consistent with the principle of low friction incremental refactoring, using a single-shared service for issuing IDs mitigates legal artifact identity issues. Each firm or legal entity could keep its own ID system but would also insert foreign keys generated from the service into to the current record store. Theses keys would be consumed via a standard based RESTful interface. This is advantageous as most modern systems and programming languages support this means of remote data/procedure access. In addition, the API layer abstracts the details of blockchain access thereby mitigating the need for blockchain enabled programmers as the majority of web and enterprise programmers are already versed in its application.

Database design is an extremely well-studied and understood discipline that's been battle tested by tens of thousands of companies. The use of primary and foreign keys to tie related data together is a cornerstone of database theory and practice. One obstacle we face is that as a global utility, the benefit of our proposal only grows as the number of users grow. Therefore, the initial development and testing will be well understood and fairly easy, predictable, and inexpensive. Selectively adding foreign keys to existing data stores offers an ideal solution because it can be tested in isolation even within a single organization's infrastructure.

The second point of the triangle is proof of existence of digital assets. In a wide range of business and legal contexts, establishing provenance and chain of custody of digital documents and data is critical. It is technically possible to store encrypted documents on the blockchain but this is not practical for a few reasons. One issue is data size. Blockchain storage is still a fairly expensive operation and the synchronization of large documents is still problematic due to current bottlenecks in network traffic and data verification/consensus. In addition, current commercially available encryption technology has a short potential shelf life. Given that the blockchain is immutable it is likely that the encryption will be broken while the data is still available on the ledger to parties other than the owner. We therefor propose that the ledger store only document hashes, aka a digital fingerprint, of the documents. A hash is the short output of a one-way function that turns any digital file of any length into a fixed size text string.

These functions, which often involve prime number factorization and elliptical curve math, have a few key features. A given input such as a binary representation of a PDF contract will always yield the exact output string, but there is no known way to reproduce or guess at the original document from the hash. If even a single character is modified in the original document, the hash of the new version will be different in a completely deterministic but unpredictable fashion. Of critical importance, even with the advent of quantum computers it is virtually impossible to reverse engineer the hash any time in the foreseeable future.

The application of hashing functions is also used to protect the blockchain itself through variations of the concept of a Merkle tree. Once data is written, it is prohibitively expense to change.

From a practical standpoint, imagine two parties digitally signed a PDF contract and recorded the hash on the Integra ledger. At some time in the future (possibly years later), the two parties have a dispute where each claim the PDF they produced is the final copy signed. This could result from fraudulent activities or simply document mismanagement. In any case, it is important to establish which, if either, is the most recently executed copy. To reconcile the situation, one would just need to hash each document and check to see if either was written to the ledger to establish which was legitimate. Even if both copies were legitimate versions, the fact that the ledger is timestamped would establish which was executed most recently. While manual reconciliation is powerful, the greatest benefit of the use of the IDs and related hashes might be document and process automation.

On the surface, the requirement for each party to maintain their own copies of the documents may be seen as a weakness. However, it follows the single responsibility principle that each component of a system focuses on a narrow function performed well. By uncoupling document storage from proof of existence, the end users are free to utilize a spectrum of different document management systems from simple local fire storage to enterprise solutions, such as Box, SharePoint, NetDocuments and so on. To reiterate an earlier point, the use of a RESTful interface makes integration with any of these systems fairly straight forward.

Consistent with the philosophy that adoption requires gradual incremental change and a focus on gaps in existing solutions, the third point of the triangle is to tackle enterprise security. More specifically, we wanted to conquer the all too common scenario where a password unprotected PDF contract is attached to an unencrypted email. There are solutions available, but they are often not used because of previously mentioned issues including fear of vendor lock-in, overly complicated solutions, and lack of standards.

The use of PGP (Pretty Good Privacy) offers a good testbed. Released in 1991, it follows an open standard and supports encrypting, decrypting and signing documents and other text artifacts such as emails. It is a good potential solution but is not as widely used as it suffers from several technical and social issues.

The first issue is that PGP utilizes a series of Public/Private Key pairs. The private keys need to be kept secret and secure, but the public keys need to be bound to a user through a unique public identifier, such as email. The issue of private key management is complex. Enterprises, such as large law firms or corporate legal departments, already utilize specialized enterprise-grade key management systems. Therefore, this is an advantage since the use of blockchain itself usually requires key management. To help smaller groups or individual users with digital keys, Integra created a local desktop application, the Integra Hub. This serves many functions including key generation and management.

The PGP public key distribution is also an issue. Public keys are generally available to your network of users. This is a challenge if this includes people outside your organization, especially if they are not well known to you. The most common solution is to publish your public key to a secure, publicly available key server, which can be hosted by MIT, Ubuntu, and PGP.org. While these are reasonably secure, there is a common fear that a hacker could execute a “man in the middle” attack, where they intersect and fake a public key lookup. Once done, they could potentially interact and decrypt sensitive emails and impersonate the other party. These fears are often blown out of proportion by the media and may not be fully justified but result in a social concern.

Blockchain can potentially mitigate this issue. We propose utilizing the Integra ledger as a PGP public key registry. As a node running on the ledger, a company will have a secured synchronized copy of the data sitting behind their firewall. As a result, a man in the middle or similar attack becomes impossible. Even if the company runs their node in the cloud or uses the API access, they will have already secured this channel to meet their security requirements.

The last issue is the PGP interface, which is often clunky. Lawyers traditionally resist change in workflow and are slow to adopt new technology. Our approach is to modify their existing tools in a fashion that is a minimal learning curve. The most obvious candidate is to leverage the new Microsoft Plugin architecture. This allows for custom HTML/Web standards plugins to be added into all office suite products and works on platforms including traditional Mac and PC desktops, mobile devices, and Office 365 enabled web browsers.

These Plugins are effectively Iframes running a standard web-based application, which allows them to interact with the RESTful interface of the Integra ledger through the simple application of standard AJAX calls. In addition, Microsoft has supplied a common JavaScript API for directly manipulating Office documents, including Outlook emails. In addition, you can create a standard CSS based interface using standard libraries, such as bootstrap, to create simple UIs. An example use case is that the custom panel can offer a button to encrypt the email. Clicking the button would make a secure HTTPs RESTful web call within the local network to obtain the email recipients public key and add PGP encryption to the body of the email so that only the recipient can decrypt.

A similar process can be used to add blockchain-based version control to MS Word. The hash of the version is stored on the ledger and the common foreign key is embedded in the metadata of the document.

When we started the initial work on the Integra ledger in mid-2017, we concluded that the blockchain substrate must be permissioned. We experimented with forks of Ethereum, such as Quorum, but ran into issues that ultimately turned us toward Hyperledger. While we have continued to keep an eye on Sawtooth and Iroha, Hyperledger Fabric was close to the 1.0 release (Currently 1.1), so Integra Inc. chose that technology. Consistent with our belief of utilizing a RESTful abstraction layer we adopted Hyperledger composer for business log (Chain Code) and RESTful interface generation.

Hyperledger Fabric works well. However, the blockchain industry, particularly Hyperledger projects, are moving towards consolidation or at least standard interledger compatibility. This is well demonstrated with the Hyperledger support for Solidity (Ethereum compatible) contracts running on Sawtooth through the Hyperledger Burrow code. As our API layers offer abstraction to the underling substrate, we have been moving towards a blockchain agnostic approach and have been experimenting with Ethereum and Sawtooth alternative DLT (Distributed Ledger Technology) backends.

The blockchain agnostic approach is facilitated by our adherence to an initial sparse API. By utilizing GUIDs for identity (Globally Unique Identifiers such as hashes, email addresses, and PGP key Strings as the main data), a developer could easily port all data to an alternate system. This is true for anyone running a node, where they have locally synched copies of all historical data. This relieves the fear of “Vendor Lock in” a major factor in enterprise utilization. The nature of the selected data also helps minimize the industry plaguing issue of scalability due to slow transaction rates.

In conclusion, there is a common central theme in this approach. Establishing a network of trust. The use of cryptographic keys drive most of this infrastructure. In the case of the Integra Ledger, we follow a consortium model. The Global Legal Blockchain Consortium (GLBC) consists of over 30 members and acts as the root of the trust chain. The GLBC grants Integra Inc. the responsibility to operate the network and run the common certificate authorities.

Ultimately, it is the GLBC and its internationally diverse Fortune 500 companies, AmLaw 200 firms, and top education institutions that protect the data in a way that no single organization can.