Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/6181--19-9967

Karle Fellowship Final Report: Development of a Chemical Blockchain with Biased Detection of Matched Taggants

Adam C. Knapp

Navy Technology Center for Safety and Survivability Branch
Chemistry Division

February 14, 2020

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
14-02-2020	NRL Memorandum Report	

4. TITLE AND SUBTITLE

Karle Fellowship Final Report: Development of a Chemical Blockchain with Biased Detection of Matched Taggants

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Adam C. Knapp

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
N2Q6

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/6181--19-9967

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

10. SPONSOR / MONITOR'S ACRONYM(S)

NRL/NISE

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

While cryptoledgers, the central novelty of blockchain technology, offer the possibility of decentralized, remotely recorded, and trustworthy enforcement and tracking systems, pure digital instantiations fail to fully address the physicality of many applications as well as the need for embedded information in the physical good itself. A possible approach when applying cryptoledgers to physical items has been to include a taggant with the item and to use this presumably uncounterfeitable tag as a physical reference for a digital cryptoledger. Such a scheme requires truthful, competent reporting to the cryptoledger across multiple transactions and presents few opportunities and no physical markings or verifications that the integrity of the tagged item is being maintained across all transaction levels. This report proposes and explores theoretical methodology for supporting chemo-physical solutions to this problem, specifically, technology from a diverse array of sources that might be adapted and used to create a cryptoledger scheme capable of operating independently of standard computational and networked protocols via chemo-physical technology, but nonetheless capable of supporting some digital component to enable faster communication and two-factor authentication for government logistics hygiene and regulatory enforcement.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Adam C. Knapp
Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	49	**19b. TELEPHONE NUMBER** *(include area code)* (202) 404-5487

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

CONTENTS

EXECUTIVE SUMMARY

While cryptoledgers, the central novelty of blockchain technology, offer the possibility of decentralized, remotely recorded, and trustworthy enforcement and tracking systems, pure digital instantiations fail to fully address the physicality of many applications as well as the need for embedded information in the physical good itself. A possible approach when applying cryptoledgers to physical items has been to include a taggant with the item and to use this presumably uncounterfeitable tag as a physical reference for a digital cryptoledger. Such a scheme requires truthful, competent reporting to the cryptoledger across multiple transactions and presents few opportunities and no physical markings or verifications that the integrity of the tagged item is being maintained across all transaction levels. This report proposes and explores theoretical methodology for supporting chemo-physical solutions to this problem. Specifically, this report explores technology from a diverse array of sources that might be adapted and used to create a cryptoledger scheme capable of operating independently of standard computational and networked protocols via chemo-physical technology, but nonetheless capable of supporting some digital component to enable faster communication and two-factor authentication for government logistics hygiene and regulatory enforcement.

This report represents the culmination of a year-long research program supported by a Jerome and Isabella Karle Distinguished Scholar Fellowship from the U.S. Naval Research Laboratory as well as the final requirement of that same fellowship. The report itself is comprised of 6 chapters: 1. an introduction to cryptoledgers and the blockchain, 2. a technical review of that technology, 3. a review of current chemo-physical taggant and anti-counterfeiting technology, 4. encryption, communication, and privacy schemes capable of non-digital physical implementation, 5. proposals for physical implementations, and 6. conclusions and futures directions. The research report is followed by appendices which document the original research proposal, other research directions related to fellowship as well as patent disclosure(s) for inventions originated by the research supported by this fellowship.

This page intentionally left blank.

KARLE FELLOWSHIP FINAL REPORT: DEVELOPMENT OF A CHEMICAL BLOCKCHAIN WITH BIASED DETECTION OF MATCHED TAGGANTS

1. INTRODUCTION TO CRYPTOLEDGERS, AND BLOCKCHAINS TECHNOLOGIES

1.1 Introduction

Innovations originating in the digital realm potentially offer new solutions like cryptoledgers and smart contracts for settling complicated multi-party transactions in a trustless, but verifiable w a y. Unfortunately, given the current arrangement of their constituent features, many of these technologies have often seemed to be solutions in search of problems. Many, if not most, transactions in government, business, and, indeed, life are neither anonymous, nor trustless, making cryptoledgers and BT less necessary than many of their proponents claim and possibly even encumbering. Further complicating the widespread adoption of these technologies are the areas of commerce in which cryptoledgers/smart contracts/BT have found application: speculative cryptocurrencies and tax evasion, and more generally, the anonymous and trustless empowerment of various facets of the informal economy. These associations have, by proxy, left significant concerns about the viability of technologies like the cryptoledger in more mainstream settings. Nonetheless, by their very success in settings of limited trust, otherwise unenforceable agreements, and few social norms, these technologies have shown promising capabilities in providing a sort of automated enforcement of informal law at very limited cost to the constituencies subject to that informal law. In doing so, cryptoledgers and their technical relations provide the possibility of a sort of governance of last resort. Thus, one could argue that this class of trust enforcement technologies are not truly technologies for business, but are instead technologies for government.

Under the auspices of that observation, the appearance of BT and cryptoledgers come at an auspicious time: Today, government agencies, like U. S. Customs and Border Protection, the IRS (Internal Revenue Service), and the DLA (Defense Logistics Agency), are being asked to do more with less in increasingly complex operating environments. Frequently, these agencies have complicated missions taking place in extended social and business networks subject to overlapping jurisdictions and regulations. Unsurprisingly, given this complexity, resource constraints and heterogeneous operating environments often force the government to make difficult enforcement and regulatory decisions.

In the context of taxation, these challenges lead to some taxes, either individually or categorically, not being collected, which encourages games of chicken between taxpayers and tax collectors and can leave compliant taxpayers to view the system as rigged or inherently unfair. Likewise enforcing accountability and honesty in supply chains is a difficult problem; the problem of counterfeit materiel can have significant and negative impact on governmental operations. For perspective as to how one governmental organization deals with the problem of supply chain hygiene, the U.S. Department of Navy "requires DON activities to implement a risk-based approach to identify and prevent the introduction of materiel that is at high risk of counterfeiting [1]" and recommends the following 8 guidelines for minimizing the operational impact of counterfeit materials [1]:

Manuscript approved February 14, 2020.

1. Purchase materiel from OMs and their authorized suppliers whenever possible. Materiel purchased from unauthorized suppliers is considerably more at risk of being counterfeit.

2. Practice proactive Diminishing Manufacturing Sources and Material Shortages (DMSMS) management. Obsolescence is a justifiable reason to purchase from an unauthorized supplier, if no other options exist. Proactive DMSMS management and technology refresh/insertion planning reduces the risk that obsolete parts must be procured from unauthorized suppliers.

3. Aggressively manage the supply chain to ensure unauthorized suppliers have been thoroughly vetted to reduce the risk of receiving counterfeit materiel.

4. Establish a risk-based set of inspections and tests proven to detect counterfeit materiel.

5. Establish a standardized process for reporting suspect counterfeit parts to all pertinent stakeholders, including Naval Criminal Investigative Service (NCIS), the Navy Assistant General Counsel Acquisition Integrity Office, the contracting officer, the pertinent chain of command (including security officer), and all users of the materiel. Never contact the supplier of the materiel. Initiate Product Quality Deficiency Reports (PQDRs) using Detailed Cause Code 5AS for counterfeit and suspect counterfeit materiel.

6. Report counterfeit and suspect counterfeit materiel to the Government-Industry Data Exchange Program (GIDEP) within 60 days of suspicion the materiel is counterfeit.

7. Train all affected personnel (e.g., program management, purchasing, inspection, test, production, engineering, quality, and repair) in the prevention, detection, containment, reporting, and disposition of counterfeit materiel, to be in alignment with DON requirements to mitigate risk in the supply chain.

8. Contractually obligate contractors and their sub-contractors to implement counterfeit mitigation practices, including those described above.

While these recommendations provide rigorous guidance for maintaining vigilance in the presence of possible counterfeit materiel, they offer few measures, other than rigorous inspection and spot-checking, to prevent the introduction of counterfeit materiel to the supply chain. They also suggest the possible efficiencies to be reaped by the introduction of a low-cost, labor un-intensive technology for regulatory verification and enforcement.

The goal of this report and the author's work effort which was supported by an NRL Karle Fellowship is to provide a theoretical underpinning for physically implementing the blockchain technology, i.e. a cryptoledger, in order to support the security of DON and other governmental supply chains as well as complex regulatory and compliance enforcement environments for physical goods. It is the assessment of the author that while digital chains of custody are relatively well covered by existing cryptoledger and blockchain technologies and research efforts, physical corollaries capable of solving the "physical last mile problem" (i.e. when a good is dissociated entirely from a digital realm) are lacking and an investigation of this subject is in order.

This report is divided into 6 chapters and accompanying appendices. The chapters cover: an introduction and overview of the underlying technologies and solutions for a physical blockchain and cryptoledger, a technical review of extant digital blockchains and cryptoledgers, a technical review of extant chemo-physical taggants and anti-counterfeiting technology, design sources for implementing physical and homomorphic cryptography, approaches to intermittent communication, methods for structured access, and mechanism

design, and finally, conclusions and final directions. The appendices document a range of associated actions and results inspired by, but separate from, this project.

1.2 Reflections on the Conception and Execution of the Karle Fellowship Proposal

The work supported by this fellowship was originally conceived with the expectation that a great deal of theoretical work would be required to develop methods capable of supporting a chemo-physical blockchain. In particular, biased sensing schemes were originally thought by this author to be a necessity to implement authentication schemes. The author has been pleasantly surprised by the intensity and quantity of research within the computer science literature during the time that this fellowship proposal was being put together and approved and subsequently executed.

The breadth, depth, and quality of the ongoing work within the field of "crypto-technology" has truly stunned and challenged the author. Due to the anticipated impact of the Internet of Things, electrical engineering and computer science researchers have worked fast and diligently to remedy the many problems that enacting a cryptoledger, smart contract, or blockchain scheme physically would entail. There has been incredible recognition and innovation in areas the author had considered requiring critical innovation for the proposed schemes to work among them: authentication with low energy needs, updateable cryptography schemes able to handle a dynamic and encrypted ledger, and the challenges of intermittent communication for such schemes.

As advanced as the electrical engineering and computer science has become with regard to the needs of this project, all of the work that the author has encountered has ultimately presumed a digital operating space. Even in the case of the internet of things, a persistent internet connection with the possibility of continual communication has been assumed to be present. Nowhere in the literature has the author found a system able to tackle "The Physical Verification Problem" or in analogy to issues associated with home delivery, "The Last Mile Problem." The development of a theoretical methodology by adapting and synthesizing these many advances has been the most important work performed by the author in support of this Karle Fellowship project. If this project were to find funding and be physically implemented, the supporting technology, as it stands today, might have the feeling of being "held together with string and sealing wax." It would, nonetheless, be viable as a working prototype from the standpoint of theoretical methodology.

2. A GENERAL TECHNICAL OVERVIEW OF CRYPTOLEDGER AND BLOCKCHAIN TECHNOLOGY

2.1 The Mechanics of Bitcoin: A Foundational Approach

Bitcoin is the first widely used and decentralized "currency" without government-backing or centralized administration via a central bank. Unlike government fiat, Bitcoin (and other cryptocurrencies) derives its worth from the belief of its users that it is inherently valuable or that someone else will readily accept it as a store of value. Its mechanics are class-defining for cryptocurrencies and cryptoledgers via the blockchain technology. The Bitcoin concept was first proposed by the likely pseudonymous Satoshi Nakamoto in a white paper entitled "*Bitcoin: A Peer-to-Peer Electronic Cash System*" [2] and first introduced as working entity by mining the Genesis (first) block on January 3, 2009 [3].

The underlying technology and mechanics of Bitcoin are defined by a public ledger which records Bitcoin transactions and is archived in a supporting blockchain. The blockchain, which contains of the transactions associated with the cryptoledger, is available for public perusal on the internet as a means to prevent

"double-spending," i.e. the spending of the same Bitcoin twice. The construction of this blockchain is a chain of blocks, where each block has a hash (cryptographic mapping) of the previous block which is used to determine the legitimacy of that block. Bitcoin addresses in the cryptoledger record Bitcoin values and are generated by random private keys, which must be kept private since like a bearer bond the control of that private key indicates ownership of the underlying asset, in this case Bitcoins. There is no recourse to a user if this private key is lost or stolen; the bitcoins addressed to it are gone. Since they are the public half of a public-private cryptographic pair, these addresses may be published without concern for compromising the Bitcoins indicated to exist by that address. This is useful for verifying that a specific user has enough Bitcoins for an agreed upon transaction if they indicate a specific address is theirs. When a user desires to spend a bitcoin, their private key is used to digitally sign the transaction, which in turn is verified by a network of nodes and added to a block which is incorporated into a blockchain.

This blockchain is maintained by that same network of nodes or "miners" which act as creators (miners) of new Bitcoins and as validators of new transactions. When a transaction such as *payee X receives Y Bitcoins from payer Z* is sent or "broadcast" to this node network, the nodes add this transaction to their respective copies of the cryptoledger and broadcast those updates to the other nodes in the network. Independent verification of a transaction is achieved when every network node has accepted a given transaction. It is in this acceptance process or *mining* where the Bitcoin blockchain reveals its true innovation.

Mining, the creation of new Bitcoins or validation of transactions on the Bitcoin blockchain, is performed using a methodology referred to as "proof-of-work." This scheme entails using processing power to maintain the consistency and validity of the blockchain across node network. Encoded in every block is a SHA-256 cryptographic hash of the previous block. *Miners*, those who perform bitcoin mining, attempt to find a number referred to as nonce so that when a block's content is hashed with that nonce, a smaller number than the network's difficulty target is found. Proving the correctness of this nonce is easy for any node, but finding such a number is computationally challenging since many different nonce values must be attempted before a satisfactory one is found. This search is underlying reason for the term proof-of-work. The computation of these nonces are incentivized by the generation of new Bitcoins, a feature of the Bitcoin blockchain which is expected to disappear by 2140, the number of which halve every 4 years. The inclusion of specific bitcoin transactions into blocks may be incentivized by including a secondary block payment to the miners with in the transaction.

It is the central conceit of the entire Bitcoin blockchain/cryptoledger/mining system that this proof of work construct is difficult computationally, that all major verifiers use proof-of-work schemes that do not give exponential hashing speed advantage over the rest of the blockchain node network, and that computing power is distributed relatively evenly amongst miners so that no cabal of miners holds the majority of computing power. Since no group has a preponderance of computing power, no erroneous transactions can be recorded and forced on the rest of the blockchain users.

2.2 Problems with the Blockchain and Subsequent Advances in Cryptoledger Technology and Applications

In the previous section, the Bitcoin cryptocurrency and blockchain was discussed and its mechanics explored at a high-level. Left unexplored were the implications of wide-spread use of Bitcoin from physical, computational/structural, and end-user perspectives.

From a physical perspective because a successful Bitcoin blockchain necessarily requires significant proof-of-work computations to accomplish its task, much physical energy (about 0.3% of the world's electrical output [4])is expended on what is an otherwise worthless computation. Given the competitive nature of mining, increasing Bitcoin prices relative to other currencies, as well as increased awareness among the general population and interest from the investing class, the amount of energy expended maintaining the blockchain is likely to grow.

Additionally, the Bitcoin blockchain has structural problems which inhibit its long term use as a medium of exchange. The average length of time to create a new block in the blockchain is about 10 minutes and is estimated to be only able to handle 7 transactions per second [5] and average wait time of 78 minutes for a transaction to be confirmed [6]. For a means of exchange with stated aspirations of competing with government-backed currencies on a global scale, the low number of transaction confirmations per second and the long wait time for confirmation present severe limitations for Bitcoin to be widely accepted. For those actually confirming Bitcoin transactions because PoW is a winner take all scheme on a per-block basis, miners are heavily incentivized to band together in mining pools where risks and rewards may be shared [7, 8] which significantly reduces the intended diversification of confirmation of the Bitcoin blockchain. In fact, over 60% of the Bitcoin mining hashrate is run by just 3 mining pools with the largest BitMain controlling nearly 51% of the Bitcoin hashrate [9] with an ever upwardly adjusting difficulty to perform those hashes [10]. This concentration represents an oligopolistic threat to the original solution to the double spending problem if those in charge of those three mining pools wanted agree to confirm incorrect transactions they could. Price fixing conspiracies, like that which occurred in the lysine market in the mid-1990s [11], have been successful among many more oligopolistic competitors and partners. On the level of individual users, such slowdowns with transaction confirmation make very real the possibility of just trading addresses for small standardized amounts of Bitcoin which are held by an escrow service or similar third party. If users were to do this, they would essentially recreate the current commercial banking system without any of the regulations and protections currently in place.

Finally, from the standpoint of the end user, Bitcoin is cumbersome to use. It requires more specialized knowledge than that possessed by the general public to get a Bitcoin wallet, find someone with Bitcoins to sell, transact on those Bitcoins [12]. Moreover, the Bitcoin blockchain's very immutability prevents the correction of human error, a sort of inevitable "last mile" adjacent problem. Most people in their daily lives as well as in the business community are used to being able to intelligently unwind and work through mistakes particularly in critical settings. Also troubling is the blockchain's haphazard record of ensuring the privacy of its users despite early claims by proponents of the total anonymity of the Bitcoin system since the ledger itself is open maintaining privacy is critical to enabling normal business operations. With Bitcoin, it remains to be seen what happens when the amount of Bitcoin generated by mining operations ceases to be of a quantity valuable to speculators. Finally, from a practical standpoint, few vendors in the world who are not directly associated with businesses surrounding Bitcoin actually accept Bitcoin as a form of payment, a situation that is unlikely to change in the foreseeable future due to the volatility of this asset. All of these defects in the Bitcoin/cryptocurrency paradigm leave end users, even early adopters, with a trust deficit with regard to this technology.

Nonetheless, the cryptoledger/blockchain/smart contract community is aware of these problems and has expended a great deal of effort to directly address these challenges. Developers have sought to tackle the most wasteful and inefficient aspects of blockchain via new methodology for handling these distributed authentication and cryptoledger systems. The developments pertinent to this report will be reviewed in a subsequent section detailing the mechanisms of the chemo-physical cryptoledger.

2.3 The Internet of Things: A Paradigm Breaking Application

Among recent proposals for new avenues and applications of the blockchain is the so-called "Internet of Things" (IoT). The IoT postulates that many common devices in private, professional, and commercial spaces should be wired and enabled for interactive support via the internet. If this vision were enacted, IoT could provide a social revolution caused by pervasive customization in every sector of society; it also enables the possibility of widespread surveillance, pervasive advertising and harassment, and low to moderate grade irritation if poorly implemented. Consequently, the specifics of the underlying technology supporting IoT are very important as the specifics of their structure and implementation define the quality of life stemming intentionally and unintentionally from IoT. In many ways, the concerns surrounding IoT, particularly when combined with a cryptoledger-based management system, mirror those associated with a chemo-physical cryptoledger as many of the concerns about physical viability, privacy, decentralization, and scale are similar.

Centralization in the management of IoT-enabled products has been a major concern and subject research amongst academics [13]. Concerns about costs surrounding maintaining computing services associated with one-time purchases and potentially infrequently used items make centralization of these computing resources problematic from a service-side business perspective as do concerns about software maintenance and updates. From an end-user perspective centralization in IoT presents all sorts of challenges concerning privacy, anonymity, and the truthfulness of service providers. In particular, for many businesses and more security/privacy-focused private citizens, the issue of closed-source code is a major one when it interacts with service providers having access to sensitive data.

However this is not to suggest that centralization is without advantages. From a physical standpoint, the IoT presents unique challenges from to decentralized computing methodologies like cryptoledgers as many of the methods developed for verifying digital cryptoledgers and blockchain-adjacent technologies do not translate particularly well to the IoT context. Physical systems that might be relevant to IoT range from across orders of magnitude from the 10s to possibly the 10^9s. Providing authentications for cryptoledgers for small networks are a completely different problem than for large scale networks like Bitcoin. For small networks, the possibility of a large scale computational attack on an otherwise small scale network are a significant possibility. Moreover, issues of intermittent connectivity with bad or partial communications is a significant problem in a potentially loosely networked physical system. All of the problems of the traditional blockchain technology like its rigidity and intolerance of error become more so in IoT. Imagine someone receiving a bad transmission for a program associated with a key fob to a rental car. In a traditional blockchain context, that person would be without remedy as that now "bricked" key fob is unusable. From the standpoint of an end-user/customer, this is a problem. Another problem for the blockchain in the IoT is that its authentication schemes are designed for large scale networks: the PoW paradigm will not work for small IoT networks since there aren't necessarily many participants in these networks, nor will the small scale computers embedded in IoT enabled devices be able to dedicate significant resources to solving PoW-type hash problems [14].

2.4 Authentication and Consensus for Digital and Internet of Things Cryptoledgers

Multiple authentication schemes, also consensus algorithms, have been proposed and their usage explored for digital and IoT cryptoledgers [14]: proof-of-work (PoW), proof-of-stake (PoS), proof-of-activity (POA), and proof-of-authentication (PoAh). As the problems with the PoW scheme have become clearer to designers and users of cryptography-enabled decentralized systems through the rise of Bitcoin, new authentication and cryptoledger schemes have emerged. Among the first alternative schemes to emerge was PoS,

in which, authentication is enabled by PoW schemes that are weighted by the significance of their solvers stake in the underlying system subject to authentication. Subsequently, PoS schemes have themselves had issues with hoarding due to the importance of seniority and stake size in determining authentication priority, which led to PoA schemes which apportion significance of authenticity due to stake size, length of use, and activity (i.e. spending and accepting the supported cryptocurrency). PoAh schemes arose as a means to achieve consensus in IoT environment without having to perform computationally and energetically costly inverse hash searches by having some authentication nodes be more trusted than others. While these are the authenticaion protocols considered below many more are available with a helpful list of their varieties be detailed by [15]. In general though according to the guidance of the Ethereum working group , all of these authentication schemes are guided by the following assumptions concerning the capabilities of cryptocurrency/cryptoledger/blockchain supporting community [16]:

> In traditional applied cryptography, security assumptions tend to look something like this:
>
> 1. No one can do more than 279 computational steps
>
> 2. Factoring is hard (i.e. superpolynomial)
>
> 3. Taking nth roots modulo composites is hard
>
> 4. The elliptic curve discrete logarithm problem cannot be solved faster than in 2n/2 time
>
> In cryptoeconomics, on the other hand, the basic security assumptions that we depend on are, alongside the cryptographic assumptions, roughly the following:
>
> 1. No set of individuals that control more than 25% of all computational resources is capable of colluding
>
> 2. No set of individuals that control more than 25% of all money is capable of colluding
>
> 3. The amount of computation of a certain proof of work function that can be accomplished with a given amount of money is not superlinear beyond a point which is reasonably low
>
> 4. There exists a non-negligible number of altruists and a non-negligible number of crazies or political opponents of the system, and the majority of users can be reasonably modeled as being close to economically rational
>
> 5. The number of users of a system is large, and users can appear or disappear at any time, although at least some users are persistent
>
> 6. Censorship is impossible, and any two nodes can send messages to each other relatively quickly.
>
> 7. It is trivial to generate a very large number of IP addresses, and one can purchase an unlimited amount of network bandwidth
>
> 8. Many users are anonymous, so negative reputations and debts are close to unenforceable
>
> There will also be additional security assumptions specific to certain problems. Thus, quite often it will not even be possible to definitively say that a certain protocol is secure or insecure or that a certain problem has been solved. Rather, it will be necessary to create solutions that are optimized for particular empirical and social realities, and continue further and further optimizing them over time.

PoW authentication systems as have been covered earlier in 2.1 are means of authentication/consensus determination which utilize a cryptographic hash as an adjustably hard one-way function. The adjustably

hard hash is intended to make the authentication process random or nearly random in the sense of who validates each block on the blockchain as any miner could do it first or be expected to maintain a preponderance of computing resources over the long term. This means that there is no incentive to validate blocks incorrectly.

PoS (coinstake) authentication schemes have slightly altered blocks in comparison to pure PoW schemes. In this scheme, the authenticator of the next block on the PoS blockchain is determined by some combination of a random input and size of wealth or length of currency ownership determined by an ever increasing quantity called "coinage" [14]. This methodology has the advantage of being low energy and computationally un-intensive to maintain. However, it, like PoW schemes, needs a large network of users to be viable in the long term and has issues with lacking distinct punishment mechanisms for creating false blockchains [17],

PoA schemes attempt to improve upon the challenges of PoS schemes by addressing issues associated with intermittent block mining, which particularly in small networks can cause fluctuations in the size of the network present to authenticate blocks as well as the natural tendency towards hoarding of currency that PoS schemes can cause. They accomplish this improvement by hybridizing the PoS scheme with the PoW scheme: Authentication starts with a cryptographic hash as in a PoW scheme, but after the hash is found, the system becomes a PoS scheme and a random set of validators derived randomly and weighted by activity and stake are then authorized to sign off on this new block. Once all of them have signed off, the mining/validation rewards are alloted and the process repeats itself.

PoAh schemes have been developed [18] specifically for the IoT in which computation and energy resources are limited, many of the node members may be inaccessible due to communication outages/blackouts, and network membership sizes are potentially small and subject to significant fluctuation in numbers. These schemes blend approaches from a variety of sources and use the user network itself to verify and maintain the underlying cryptoledger, which is no longer stored in a block*chain* but in a directed acyclic graph, which enables a decentralized and asynchronous verification process for parallel transactions. While this seemingly solves all of the problems of the blockchain, the PoAh requires an initial authentication step to be allowed into the network in the first case meaning a gatekeeper to entry is required, which works well for IoT "walled garden"-type applications, but is not appropriate for cryptocurrencies.

3. CHEMO-PHYSICAL TAGGANT AND ANTI-COUNTERFEITING TECHNOLOGY

3.1 An Introduction to Chemo-Physical Taggant and Anti-Counterfeiting Technology and Applications

Technologically separate from our discussion of digital crypto-enabled technology in the previous chapter are chemo-physical taggant systems and anti-counterfeiting technology. Unlike, our earlier discussions taggants and anti-counterfeiting technologies have far more modest goals than the blockchain technology. In general, taggants are used to mark a physical object as belonging to or being associated with someone, some group, or organization. Classic examples of taggants are used for theft deterrence by marking an item as being owned by or originating from a particular source or entity. Likewise this same technology may be used to track a specific technology as with Swiss explosive taggants. In cases where authentication of source and quality is important as in pharmaceutical distribution taggants are widely used to prevent counterfeiting. They are also used for regulatory and tax enforcement by marking fuel as being compliant with specific aspects of the tax code in various countries.

In general a good chemo-physical taggant is difficult to duplicate, is resistant to the elements and to acts of God, and are easy to detect and trace in the event that a tagged good needs to be verified for some reason. Often for chemical taggants this means a spectroscopic measurement as such a measurement may be performed quickly and non-destructively with simple equipment. When used for forensic type purposes taggants are effectively used for four purposes as originally described in Gooch et. al. [19]:

1. **Monitoring**: Taggants may be affixed to exteriors and entry points of protected areas and objects in order to be monitored for disturbances. Individuals and vehicles which disturb these areas are then marked and potentially trackable if the demarcating taggants are successfully transferred. In situations, where the prevention of crimes like trespassing is important, the transfer of these specific taggants provides strong physical evidence of criminal activity [20]. However, care is required that secondary erroneous transfers to innocent bystanders do not also occur as has happened in some cases with physical DNA evidence [21].

2. **Anti-Counterfeiting**: Intellectual property-dependent industries like publishing, clothing, pharmaceuticals and semiconductors as well as government document regimes like currencies, passports, personal identification cards, etc. are subject to counterfeiting activities and in cases where high quality counterfeit goods are produced are difficult to distinguish from the authentic article. Taggant technology and marking agents are often deployed to prevent counterfeit articles from being circulated as legitimate [22, 23].

3. **Tracking**: Materials, like precursors for controlled substances or the controlled substances themselves, are tagged as means of large-scale supply chain control. Instances of this occur in the production of plastic explosives where trace chemicals are added which detectable by canine law enforcement units [24]. The covert inclusion of taggants into illicit substances has also been proposed as a means of mapping supply chains [25].

4. **Property Marking**: Valuable items may also be intentionally tagged as a means of indicating ownership. Often these taggants have a specifiable combinatorial factor or "bar code"-like feature which allows a tagged good to be looked up in a large database and its ownership confirmed. Technologies like SmartWater^{TM} [26] are used for this purpose as are DNA taggant technologies [27].

Explosives taggants are here described as a special separate form of taggant due to the unique applications of the tagged good. Volatile taggants are included in some explosives by international convention as a means to help prevent airline terrorism and are intended to enable canine detection prior to detonation. Currently, Switzerland is only country which requires taggants in certain classes of imported explosives that are rugged enough to survive detonation. Due to the survivability requirements of these "post-detonation" taggants, their information encoding capacity is typically reduced resulting in a limited number of tags. This combination of limited number of tags and extreme survivability present a possible persistence problem as this sort of taggant could build up broadly throughout built society and thus nullify the utility of the taggants themselves.

Fuel taggants are used by tax authorities in their respective countries and jurisdictions to mark fuel that is either taxed or subsidized differently for a specific purpose than fuel in the primary market of the subject locale. These taggants themselves are either liquids or solids which dissolve easily in the target fuel and change the color of both the fuel itself and its emissions when burned and are difficult or cost ineffective

to separate from tagged fuel. This allows for law enforcement to enforce legal fuel usage. While in theory fuel taggants could come in many different hues in most countries only one or two are actually used, so this taggant class has the lowest information encoding capacity of any chemical simulant, but the easiest means of detection, the human eye.

DNA (deoxyribonucleic acid) taggants are the most recent class of taggants explored here. Unlike the other taggants presented which are grouped by application and tuned to the context and life cycle of the tagged good. DNA taggants are as the name suggest short strands of DNA which are added, affixed, or adhered in some way to a target good. In comparison to the chemistries and physicality of most taggant systems, DNA is quite fragile. However, it makes up for this fragility with information coding capacity. It's subsection below will briefly explore the commercialization, applications, advantages and disadvantages of this particular taggant form as it is most recent "novel" chemistry to transition from the lab to the commercial world.

Counterfeiting and intellectual property theft deterrence are another major application of taggant or taggant-adjacent systems. For anti-counterfeiting efforts, taggants and related technologies are used as certificates of authenticity. Often a tagged good or currency will have both well known taggants present on such an item to allow the public to verify authenticity at a basic level as well as hidden taggants and features known only to experts and the manufacturer(s) to catch more adept counterfeiters and forgers. A brief discussion of taggants used as deterrents for counterfeit pharmaceuticals and watermarking technology will be briefly discussed.

3.1.1 Theft Deterrence

Theft deterrence is the most common reason for using chemical taggants. Commercial technologies like SmartWater^{TM} are used as forensic asset marking systems. The general intention of such taggants is to mark valuable goods with microscopic or molecular and difficult to remove taggants which can be directly related to a specific owner. There are, however, concerns by some security experts that theft deterrent taggant technology could be fraudulently used to falsely mark property not belonging to the owner [28].

Uniquely marked taggants are often batch-made in relatively large quantities by manufacturers to ensure uniformity in the unique marking capabilities of a particular run. Each run is then sold in its entirety to a specific client and the unique properties of that particularly batch are then attributable to that specific client. The specific chemo-physical relations used by these taggants vary by manufacturer and across product lines.

Gooch et. al. [19] state that "A forensic taggant displaying entirely ideal characteristics should be of low cost to produce, have high coding capacity, be non-toxic to individuals and the surrounding environment, be simplistic and inexpensive to detect and analyse via non-destructive means and be of a complex enough nature to prevent duplication." Forensic chemo-physical taggants typically come in 3 different classes of taggant/detection type: physical taggants, spectroscopic taggants, and chemical taggants.

A physical taggant is generally a durable and inert solid particle with specifiable size and structure which may be mixed with or affixed to the tagged good in question. For instance, microdot technology developed during WWII was re-purposed postwar by companies like DataDot and Microtrace as a tagging technology in which small polymer microdot disks of size $2\mu m - 1000\mu m$ are suspended in coatings like varnish, paint, or ink. These microdots which are imprinted with a code could then be collected forensically or read directly *in situ* via microscope and corroborated by the manufacturer's client-owner database. $3M$, later

MicroTrace Solutions [29], developed a durable colored polymer sandwich capable of surviving detonation events of explosives. The color sequence of the requisite layers determine the origin and ownership of the materials. While the coding capacity of such a taggant is lower due to practical limits on the number of layers and number of plastic colors, the durability of this particular class of taggants is the most important feature of this taggant [19]. This taggant may be physically collected and verified by optical microscope or magnification device [29].

Spectroscopic taggants attempt to put unique, tunable mixtures of optically active compounds like non-toxic organic on tagged goods. Spectroscopic tagging allows for high-throughput through verifying measurements via spectrophotometry. While the dyes used for these sorts of taggants are cheap and readily available, this is a proverbial "double-edged sword" as sophisticated or at chemically savvy counterfeiters are sometimes able to independently replicate them.

While Chemical taggants as defined by the forensics literature are trace additives to a tagged good whose presence or absence at a particular concentration threshold represent a binary taggant code, which is used to correlate with membership information in the manufacturer's database. Isotopic materials can also be used in this fashion although they are expensive and difficult to prepare. A typical chemical taggant is the SmartWater system which uses rare-earth materials like the lathanides to tag goods due to their rareness and the general challenges associated with synthesizing them which make them difficult to counterfeit.

3.1.2 Explosives Taggants

Taggants are used to mark and track explosives in both their pre- and post-explosion forms. Due to an international agreement, "Convention on the Marking of Plastic Explosives for the Purpose of Detection" sponsored by the International Civil Aviation Organization plastic explosives must be marked [24]. Within the USA, volatile chemical markers like 2,3-dimethyl-2,3-dinitrobutane are added as taggants for unexploded plastic explosives in accordance with the previously discussed convention.

However, it is Switzerland where explosive taggants find their biggest application due to Swiss laws and regulations requiring the tagging of explosives which are also via the information carried by their tags enmeshed with a permitting, distribution, storage, and licensing regime extending down to the Canton level [30]. Tags are applied to working/commercial explosives for uses like blasting, but usages for military, entertainment (fireworks), and sporting purposes are exempt. It is important to note that the Swiss consider the tags themselves to be part of a much larger integrated regulation regime for explosives. For instance, explosive sellers are required to buyback unused explosives which are destroyed or re-integrated into new products and all detonation/explosive control devices are tagged as well.

3.1.3 Fuel Taggants for Excise Tax Evasion Prevention

Fuel taggants for excise tax marking represent a unique class of taggants. While most taggants are physical solids during the standard used conditions of the tagged good, fuel taggants are liquids used to change the color of the marked fuel and its emissions for the purpose of dividing fuels which are functionally the same but taxed at different rates dependent on purpose into different classes. In many countries, like the United States, but also Spain, France and Poland, agricultural and marine fuel, which are essentially diesels of varying heaviness, are taxed at different rates as means to economically support agricultural and marine-based industries.

In order to prevent citizens from buying these untaxed/lower taxed fuels nominally for these purposes, but actually for other purposes like driving, chemicals are added to them which are both difficult or cost prohibitive to separate out and which change the color of the fuel and its resultant emissions when burned.

3.1.4 DNA Taggants

Deoxyribonucleic acid (DNA), the building block of the genetic code, has the advantages of being nearly non-toxic, almost unlimited coding capacity for unique identifications and relatively cheap synthesis for taggants. As a consequence, several companies like Applied DNA Sciences, TraceTag, and Selectamark Security Systems have commercialized DNA for tagging goods [19]. Unfortunately the verification of these tagged goods has some challenging aspects due to the labor and relatively capital intensive nature of sequencing the DNA taggant as well as the physically sensitive nature of the DNA itself. It remains to be seen whether or not DNA will prove to be a practical, albeit versatile, molecule for tagging [19, 31].

3.1.5 Intellectual Property Protection and Watermarks

Intellectual property protection is a major application of taggant systems in the commercial world [32, 33]. As in other taggant applications, a variety of taggant technologies are available for intellectual property theft prevention. Securalic [34] taggants seek to hide particles within materials to act as provenance indicators as is typical for taggant technology many of the internal details about these particles are not readily accessible to the general public. Likewise Applied DNA Sciences uses DNA-based tags to mark a variety of goods ranging from textiles to motor oil [35]; as in the case of Securalic, many of the details surrounding the specifics of their technology are suppressed.

3.2 Current Research in Chemical Taggant Technology

In this author's opinion there is a large gulf between the extremely practical nature of commercialized taggant technology and academia's innovative, but sometimes fanciful concepts for chemical taggants presented in the research literature. In general, commercialized taggants are designed to be as chemically or physically simple as possible. With the notable exception of DNA-based technologies, they are as robust and as simple as possible across a wide variety of relevant physical and environmental conditions. Notably, novel chemistries are avoided, so are environmentally sensitive and expensive detection requirements. Also avoided in commercialized taggant technology are detection methods which require specialized skills for interpretation.

The results of academic research for taggant technology are often completely counter to this commercial approach. Academic research for chemical taggant technology is often focused on novel chemistries and challenging detection schemes with little concern for the practical contextual and environmental realities faced by chemical taggant technology. While much of this work fails to find application, it does act as a platform for many interesting new technical ideas that have the potential to vastly expand what a chemical taggant can be as well as provide a proof-of-concept for those same new ideas. In particular, the following section tries to highlight recent advances in areas like chemical logic, chemical steganography, supramolecular taggants, and physically unclonable functions as well as provide a context for recent work mixing information and chemical systems.

3.2.1 Chemical Logic

Molecular logic seeks to implement the basic elements of computation Boolean logic gates in chemical form at the molecular scale. While researchers have come to grips with high likelihood that such approaches may never challenge classical silicon-based microprocessors, let alone nascent quantum computing technologies, they do provide an interesting avenue into placing computational power in unexpected places like pharmaceuticals for drug delivery or *in situ* sensing [36]. A simple application of molecular logic is to a molecular keypad: In a molecular keypad n 2-input AND gates are concatenated such that one of the inputs of $n-1$ of the gates is the preceding output from the prior gate. If all $n+1$ external inputs are correct then the final gate evaluates to true and the "lock" opens. In chemical terms this may be setup as reaction hierarchy where each reaction intermediary gives the desired product only if the proper compound is chosen from a list of possible inputs. This sort of chemical keypad has been synthesized and implemented using chelation agents, EDTA and light in work by Marguilies et. al. [37]. A supramolecular chemical keypad switched with light pulses was given by Adréasson et. al. [38].

3.2.2 Chemical Steganography

Classical steganography seeks to hide important secret information in an otherwise innocuous covertext. Sarkar et. al. [39] have implemented such a system and more in an intrinsically chemical medium which allows a properly equipped practitioner to conceal messages within the emission spectra of a unimolecular fluorescent sensor. Moreover, the methods of concealment enshrouding these messages are via steganography, cryptography, and password protection. This system, which is referred to as a molecular-scale messaging sensor, is able to convert randomly chosen chemical signals into emission patterns. The authors claim that its versatility as a chemical sensor allows it to hide information as well as encrypt and decrypt information via the specifics of the sensors' tuning.

3.2.3 Supramolecular Taggants

Supramolecular chemistry has proven to be a fertile ground for taggant technology due to the tunability of the chemical structures created as well as their interactions. Lathanide complexes have been commercialized as taggants due to the tunability of their uniquely hued compounds which are used to mark dates in tagged goods [40] and which have complex spectra amidst an incredibly high multiplicity of possible compounds which makes them challenging to reverse engineer and hence fraudulently produce. F. Stoddart et. al. [41] has filed a patent application for supramolecular, stimulus responsive, fluorescent dyes for security inks or taggants, with spectra that are tunable with the relative concentrations of the constituent compounds. Finally Carvalho et. al. [42] have created a supramolecular chemical keypad system in which host-guest systems are employed as the logical constituents of the keypad system.

3.2.4 Physically Unclonable Functions and Chemo-Physical Authentication Technology

Physical Unclonable Functions (PUFs) are physically-derived mathematical functions [43] which map stimuli C_i to responses R_i, are ideally invariant to external physical factors like temperature, pressure, or humidity, and which depend upon unpredictable and uncontrollable features to provide their uniqueness such as the random microstructural aberrations in a semiconductor introduced during the manufacturing process or the geometric configuration of the cracks in a polymer craze. Since these functions are essentially random and uncontrollable they may be used for authentication needs and if the randomness of the underlying physical phenomenon is understood well enough, as seeds for random number generators.

4. ENCRYPTION AND MECHANISM DESIGN FOR PHYSICAL CRYPTOLEDGER AND BLOCKCHAIN TECHNOLOGY

4.1 Introduction

Implementing a cryptoledger in a non-digital physical system requires careful consideration of the currently available and appropriate encryption technologies that can be practically adapted from the digital realm. Just as important to making a physical cryptoledger are other supporting services which have been implicitly solved in digital systems which need to be explicitly worked out like problems associated with intermittent non-digital computation, communication, and structured access to encrypted data. In the following chapter, methodology from the literature that will be used to address these problems is presented. This presentation is intended to give a first look at these techniques within their original context without the influence of the chemo-physical ledger impinging on their individual workings.

4.2 Visual Cryptography

Visual cryptography is a unique cryptographic technique in which information is encrypted as some number of noisy images on transparencies, which when some subset of those images are superimposed on top of each other reveal their encrypted content, typically unrelated de-noised images. In the context of a chemical cryptoledger, this form of cryptography is interesting because it represents a tangible, pre-existing form of cryptography with decryption steps (the stacking of images, followed by human viewing) that occur independent of digital technology. For digitally-independent or minimal chemo-physical cryptoledgers, visual cryptography is a practical, secure, and implemented form of cryptography which can be adapted from image-based to physical-based cryptography.

Arguably, the best known/archetypal example of visual cryptography was developed by Moni Naor and Adi Shamir in 1994 [44]. In this visual-sharing approach, a plaintext (target image) is encrypted as $N \geq 2$ shares so that only by possessing all N shares can the plaintext image be recovered. From a physical perspective, processing is performed on the plaintext image so that N shares are generated. These shares are then printed on transparent sheets. Overlaying any combination of these sheets with total number less than N yields no image related to the plaintext, but overlaying all N sheets reveals the original cover image.

Since 1995 many variations and improvements of visual cryptography have been published such as: k-out-of-n visual cryptography in which any subset of k shares may be used to decrypt the enciphered image [45, 46], recursive hiding of secrets so that an image is decrypted so that different parts of the image are successively revealed as more and more shares are successively superimposed on top of each other [47], size invariant visual cryptography which is a more refined form of visual cryptographic encoding [48, 49] as well as visual cryptography for greyscale, halftone, and color images [49–51].

4.3 Homomorphic Encryption for Digital Systems

Homomorphic encryption is a method of encipherment enabling computation on and with encrypted data. While current digital implementations of this encryption technology are extremely slow, researchers expect that successive improvements to this methodology will allow for faster computation enabling widespread cloud-based computing on encrypted data [52]. In the context of chemical cryptoledgers and blockchains, homomorphic encryption is a useful technology even if slow since it allows encrypted portions of the ledger to be updated. Moreover, visual cryptography-specific flavors of homomorphic encryption [53–55] have

been developed and thus this provides a smooth path for technology transition into the non-digital chemical domain. However,this subsection is a brief description of the basic mode of operation of this technology in the *digital* domain at a fairly high-level to give the reader a sufficient background in subsequent chapters and sections.

Homomorphic encryption as a subject of study was first proposed in 1978 [56], but it was not until 2009 that the first successful fully homomorphic[1] encryption scheme was developed and published by Craig Gentry [57, 58]. In general, fully homomorphic schemes work by bootstrapping a partially homomorphic[2] encryption algorithm into a fully homomorphic encryption algorithm. Gentry's key contribution was recognizing that a partially homomorphic encryption algorithm that can encrypt and decrypt itself and perform at least one more operation within correctable error is capable of being "bootstrapped" into a fully homomorphic encryption scheme by encrypting data, operating upon it, decrypting that data at this second encryption level, correcting the noise from the operations and then repeating this whole process.

In essence, Gentry realized that he could embed a small "computer" within the encrypted data and then use that small "computer" to operate on the underlying data in-between successive encryption-decryption steps followed by correction steps. This lurching (n+1)-steps forward, n-steps back process also explains the general slowness of fully homomorphic encryption: Most of the computational operations used by this fully homomorphic encryption scheme is expended on the internal encryption and decryption steps necessary to protect the security of the underlying data with a comparatively much smaller number of operations actually used on the computation of interest.

Specifically, Gentry developed a cryptosystem capable of supporting the requisite *encryption/decryption* functions as well as an *evaluate* function which acts on a description of the program being run within the cryptosystem [59]. This input program is specified not as a typical sequential program, but as a Boolean circuit or network composed of the standard Boolean functions like AND, OR, NOT, XOR, etc. and which can be specified by addition and multiplication operations. It is in this sense of a Boolean network that a small computer is instantiated within the cryptosystem. Theoretically, it is possible to run any program within this computer.

In practice, as operations are performed on the underlying enciphered data they build up discrepancies between their operational values and the values that we as users desire. These discrepancies are referred to as "noise" which while they are treated probabilistically are not truly random or pseudorandom. Successive operations on this now "noisy" data cause these discrepancies to build up; if it is not corrected for this noise or errors will overwhelm the true/desired output. In order to rectify the noise, the cryptosystem's *evaluate* function runs its *decrypt* function with an encrypted key and resets the data from its noised state. This process is then repeated until the entire desired Boolean circuit is run on the encrypted data.

4.4 Intermittent Communication: The Paxos Algorithm and Beyond

The Paxos algorithm was originally developed by Leslie Lamport [60] for enabling fault-tolerant computing on distributed systems. It is a consensus or synod algorithm which seeks to find agreement amongst distributed nodes or users in the presence of noisy and intermittent communication. Lamport defines a consensus algorithm as being an algorithm which chooses a single proposed value from a collection of value

[1] In this context, the term "fully homomorphic" means that there are no limitations on computations that may be performed on the encrypted data.

[2] A partially homomorphic encryption algorithm is able to implement all of the basal units of computation within error.

proposing processes with the caveats that if no value is proposed, no value is chosen and that the processes should be able to learn of the chosen value with following safety requirements: only a value proposed by a process may be chosen, only a single value is chosen, and a process never learns a false value. Consensus algorithms, in the form of the Paxos algorithm, are reviewed here due to the importance they play in reaching consensus in the context of cryptoledgers. Specifically, this subsection is intended to show that fault tolerance is not driven by digital communication, but by procedure. The Paxos algorithm demonstrates this assertion in its original formulation [60] by being implemented in terms of people and pieces of paper. While obviously slower than a digital system such a verification routine is still possible and practical in a physical medium subject to human intervention so long as the protocols are followed. Subsequently, the following parts of this section present the Paxos algorithm protocol without recourse to digital-specific terminology to give the reader a sense of how such an algorithm might be implemented in a physical system.

The actual implementation [61] of this algorithm is divided amongst 3 classes of agent: proposers, acceptors, and learners. The goal of these 3 agent classes is to agree on a value; this value could be anything like the remaining balance on a car loan, a price for a pair of shoes, or an identification number for a client. The value itself doesn't matter, what matters is that a collection of distributed ledgers or servers all come to the same conclusion. These agents may communicate asynchronously at an arbitrary speed, may fail by stopping, and may restart. If all agents fail after the selection of a value, then a solution is impossible unless that value is remembered by an agent after restarting. Finally, messages can be duplicated, lost, or be delayed, but they are never corrupted.

Superficially, the Paxos algorithm seeks to perform the following steps:

1. A *proposer* proposes what it thinks is the correct *value* to the *acceptors*.
2. The *acceptors* look at this proposal and either accept or reject it.
3. If a majority of the *acceptors* accept, then the *value* is chosen to be the actual *value* and is broadcast to the *learners*. [62]

However, more machinery is required to prevent collisions amongst proposers if more than one proposer have differing values. The Paxos algorithm instead follows the following steps to prevent collisions amongst proposals [62]:

1. Every proposal $\langle n, v \rangle$ should have two parts a sequentially increasing proposal number n and a proposed value v. Each additional proposal is a "retry," another attempt at an agreement about a value.
2. A prepare request is a proposal sent by a proposer to the acceptors.
3. An acceptor upon receipt of a prepare request notifies the proposer that it will not accept subsequent prepare requests with a lower proposal number.
4. On the occasion that an acceptor has already accepted proposals, that acceptor will respond to the latest proposal with the proposal that has the greatest proposal number that is less than the new proposal request. For instance if an acceptor receives the proposal $\langle 5, 10 \rangle$, but has already received the proposals $\langle 1, 20 \rangle$, $\langle 2, 30 \rangle$, and $\langle 4, 40 \rangle$ then that same acceptor will respond with $\langle 4, 40 \rangle$ and a promise not to accept proposals with proposal numbers lower than 5.

5. Once a majority of acceptors have responded to a proposer, the proposer will send an accept request to each of the acceptors with a proposal number of n. The value v in the proposal can be either a new value or the corresponding v of the highest numbered proposal.

6. An acceptor accepts an accept require only if it has not received another prepare request in the intervening time with a higher proposal number.

7. When an acceptor accepts a request, it sends it out to all of the learners that it is has accepted it. By virtue of a majority of the agents doing this, it will be the accepted value of the system.

The Paxos algorithm allows for a distributed system to agree (eventually) on any number of agreed upon values. Using blockchain technologies, it can be updated to prevent cheating schemes, but at its core it enables the establishment of consensus in a noisy transmission system.

4.5 Methods from Differential Privacy and Structured Access Methods

Areas of study that might inform future work on the chemo-physical cryptoledger are differential privacy and structured access methods. Differential privacy might become important as chemo-physical cryptoledgers enable data-sharing within and potentially outside the cryptoledger chain and the various members don't want to share the exact values of their ledgers, but see value in sharing some information so long as individual privacy is not violated. Likewise structured access methods present a more refined way of enabling and restricting access to various quantities encrypted in the cryptoledger.

Differential privacy as a field of study and a methodology seeks to place provable constraints on algorithms used to disclose information from statistical databases so that the underlying contributors and *non*-contributors to those databases remain anonymous with very high probability. Applications for these techniques occur in privacy critical situations like the release of information, tabulations and statistics from datasets like those collected by the U.S. Census Bureau [63, 64]. In the context of a chemo-physical cryptoledger system, differential privacy techniques offer the possibility of targeted and anonymized user data collection and disclosure.

Modern differential privacy was founded in a paper by Dwork et. al. [63] where they defined ε- indistinguishability and gave algorithms for satisfying its requirements. In that paper, the authors consider a statistical database, which is described as being the collection of entries associated with a representative sample from some underlying population. They then describe their work's goal as allowing a database user to "learn the properties of the population as a whole while protecting the privacy of the individual contributors." In order to protect the privacy of the database contributors, Dwork and her co-workers determine the right amount and type of noise that needs to be added to the data table such that the calculation of the statistic in question does not impinge upon the privacy of the database contributors.

Structured access methods are means of controlling a diverse hierarchy of privileges and complex interactions among a set users. This task may be understood in analogy to a commercial lock system: Imagine that one owns a multi-use, commercial building with a mix of residential, commercial, hotel, and light industrial tenants. Each of these building users (tenants, building personnel, and building owners/management) will have a variety of demands on the lock system for entire building. Some of these demands will be compartmentalized to the specific area of the tenant such as access to an apartment or office, while others, such as the access requirements of the building engineer, may range throughout the building and others, like suppliers and distributors, may need access to most or even all of the present concerns depending on ever

protean business relationships. Servicing these relationship/access needs is challenging when the users want to move beyond the multi-key, siloing model in which access is compartmentalized at some level and users with cross-cutting needs have multiple keys for specific parts of the building.

Recently, methodology has been developed by Voemel et. al. for calculating and encoding these privilege hierarchies in the context of developing commercial lock systems [65]. This is accomplished by encoding keys and locks (or lock cylinders) via a binary vector model and then analyzing the privilege hierarchy via partial ordering and upper semi-lattice structure techniques. Ultimately though, the assignment of unique keys capable of encoding privileges in an optimal way is found to be NP-hard and thus a variety of algorithms that are "close enough" to optimal are proposed. Encoding complicated privileges for structured access control is thus a difficult task for any network and not just in the context of a chemo-physical cryptoledger system.

5. CHEMICAL CRYPTOLEDGERS AND BLOCKCHAINS

Chemical cryptoledgers have been proposed as a means to physically track and label goods with respect to any sort of enforcement regime such as excise collection, regulatory/contract validation, or property verification. Implementing a cryptoledger in a physical regime that potentially has no digital aspect is appealing because it inherently makes physical custody and interaction the means of information transference and removes the possibility of the problems associated with digital hacking. If a digital element is incorporated as part of the chemo-physical cryptoledger scheme, the scheme automatically provides a form of two-factor verification via its digital and physical elements. From a practical standpoint, a chemo-physical cryptoledger also provides a low/no energy profile way of providing verification services as well as a possible means to do so without specialized detection equipment if human sensory organs and responses are used as detection devices. In this chapter, we will review potential applications and specifications for creating a chemo-physical cryptoledger with available technology as well as explore areas where potentially more work is needed.

5.1 Government Applications for Chemical Cryptoledgers

The needs of government for a chemo-physical cryptoledger differ greatly from what the private sector might require of such a system. While the private sector might be interested in enforcement with regard to contracts or property rights be they intellectual or otherwise, government and its constituent entities are often interested in regulatory enforcement which potentially touch on a variety of interacting legal requirements and regulatory regimes. Also, the privacy concerns of the government with respect to any collected data potentially differ from those of private entities and are generally magnified in comparison to the standards applied to the private sector. The government also operates in a variety of high-security regimes that are not present in the private sector. The following subsections outline 2 example government uses for a chemical cryptoledger, the first in securing a supply chain for the DOD and the second as a means of distributed taxation collection and regulatory enforcement.

5.1.1 Example: A Secured Government Supply Chain

Within the activity space of the DoD and other government agencies, there are many supply chain custody and verification scenarios that need to be addressed to address both security concerns as well as legally mandated regulatory and purchasing requirements. Consider for example a supply chain supporting the manufacture of electronic equipment, each unit of which is comprised of possibly a few hundred to a thousand components. A variety of requirements for this supply chain need to be addressed: The provenance

and custodial control of each of these components is important to ascertain and verify in their individual movements through the supply chain as are the specifics of the handling of each component at each stage of manufacture. Complicating these verifications and validations is that these components are potentially subject to assembly mid-transit through the supply chain and both the individual components as well as their assemblies are potentially dual use. These factors present While in practice it is impractical to re-verify each component at every step of the chain, if for no other reason than the components become incorporated into a greater aggregated assembly i.e. electronic components soldered to a circuit board, the aggregated whole can itself become a component with that contains pointers to it verified sub-components for re-verification if necessary. In this way, assemblages of sub-assemblies may be incorporated together until a final physically verifiable product is produced.

5.1.2 Example: Distributed Enforcement for Taxation and Regulatory Needs

Another possible application of a chemical cryptoledger is to the cross-validation of regulatory and tax compliance through multiple jurisdictions. Validating regulatory compliance like a secured supply chain, validating and presenting compliance with regulatory and tax authorities is fairly similar in essence to securing a supply chain, albeit with potentially fewer aggregation steps. Like a supply chain, an underlying cryptoledger is needed to attach compliance uniquely to a physical good. By incorporating, this readable information onto good itself without recourse to networked systems it becomes possible to operate in challenging environments.

5.2 Designing Physical Systems to Support Cryptoledgers and Blockchains

Constructing a chemo-physical cryptoledger system requires a means of communication or information transmittal, a secure cryptography protocol, a computation platform, a consensus protocol, and a means and protocol to physically ”read and write” or pattern chemicals. The following sections present a plausible path for creating a physical chemo-physical blockchain. We outline the approach here and present approaches to handle communication and information transmittal.

Currently existing methods are available and adaptable to a chemo-physical cryptoledger. In subsequent sections we will explore using the DLedger [66, 67] cryptoledger/consensus protocol as means for handling a highly decentralized cryptoledger with intermittent communication and low/no-power requirements. Our cryptography and computing platform will be supported by a homomorphic visual cryptography system capable of supporting simple Boolean function calculations

At the most basic level, communication of information may accomplished via point to point contact under the assumption that all relevant network members are still in contact with the network. This particular communication method requires no support infrastructure other than the ability to bidirectionally exchange information in any interaction.

5.2.1 Homomorphic Visual Cryptographic Approaches to Chemical Cryptoledgers and Blockchains

Multiple groups of researchers [53, 54, 68] have developed visual cryptographic schemes capable of supporting homomorphic computation which are potentially suitable for use in a chemo-physical cryptoledger scheme. This section will focus on the homomorphic visual cryptography scheme developed by D’Arco and De Prisco [68] as this is the scheme that is proposed to support a chemical cryptoledger. D’Arco and De Prisco propose combining two earlier methods garbled circuit construction and visual cryptography to

allow the secure evaluation of a bivariate function f so that it may be evaluated via "shares method" (i.e. transparency superposition) typical of visual cryptography. Just as important while this method is limited to pairs, which is sufficient for the pairwise interaction sought for chemo-physical cryptoledgers, it can evaluate arbitrary functions. This means that it can support any cryptoledger protocol, albeit probably very slowly. While the specifics of encryption for chemo-physical cryptoledgers have much room for improvement, the skeleton of a solution is available now.

5.2.2 Authentication and Consensus Algorithms for Chemical Cryptoledgers and Blockchains

Perhaps, the most challenging aspect of a physical implementation of a cryptoledger is the authentication and verification requirement of such a scheme. Bitcoin's blockchain technology presumes persistent connectivity, communication and computational resources in and available from its supporting digital network. For a physical implementation, such resources do not necesarily exist, and if they do, are present in a greatly diminished form. Also, because of the possibility of an incompletely networked system for long periods due to physical items being shipped through the supply chain or being stored in a warehouse, a physical cryptoledger network would need the ability to partition and reconstitute itself at a later date.

Fortunately, many of these problems have been considered in research on the internet of things, which while still networked has many of the same problems as the chemo-physical cryptoledger. There have been a variety of approaches to these problems in the literature, which are amply reviewed in [66]. Key innovations have been developed in the literature over the past 5 years in dealing with the various problems surrounding a cryptoledger technology adapted for objects in the physical world.

In particular, the literature has moved away from a linear descent-type blockchain cryptoledger approach to a more decentralized graph-based distributed ledgers system as seen in cryptocurrencies like Byteball [69], Nano [70], and IOTA [71]. In particular these systems have focused effort on developing cryptoledgers on decentralized acyclic graphs, which have a number of benefits such as supporting partitions. Unfortunately, some or all of these technologies still require connectivity, power, and significant computing power since IOTA and Nano require proof of work for verification and validation and Byteball has a linear main chain structure with third-party users which require persistent communication capabilities.

DLedger [18, 66, 67] is a decentralized acyclic graph-based cryptoledger technology which has been developed to remedy the previously discussed problems and to directly address the needs of the internet of things in low power, low computational resource environments with intermittent connectivity and partitionable cryptoledgers. It aims to satisfy 3 conditions [66]:

1. **Providing robust ledger whose data consensus is resilient to unstable IoT network conditions.** In the case of network partition or intermittent connectivity, entities from different subnets should still be able to contribute to the ledger system. After network failure, the distributed ledger can quickly recover from the partition by aggregating the data generated by different subnets.

2. **Working with constrained capacity of IoT devices and the massive scale of data.** The distributed ledger should be efficient enough for constrained devices to append their own data and at the same time, preventing the potential abuse and attack scenarios.

3. **Filling the gap between inefficient data dissemination in IoT and the high data throughput required by the P2P network.** The ledger should support efficient data dissemination for the routine synchronization among peers.

This DLedger scheme should be implementable in a chemo-physical cryptoledger system. By taking advantage of the secure homomorphic computing facility provided by the visual cryptographic to process the ledger's functions and provide information exchange security during physical transfers, the DLedger should be directly implementable as is although the specifics with regard to memory and speed are obviously good and taggant-detector system specific.

5.2.3 Chemical Detection Schemes to Support the Chemical Blockchain

In order to extend visual cryptographic methods to the realm of chemical and more general physical detection, it is important to consider what visual cryptography is from a physical standpoint - a 2- dimensional patterned Boolean (black and white) array determined by the superposition of 2 printed transparencies. From a computational standpoint, this task is 2-dimensional Boolean array determined by the Boolean OR-summation of two or more shares (Boolean arrays), so in essence the only information that a chemical cryptoledger needs to carry is a Boolean array which conforms to some agreed upon standard.

The actual physical encoding of that Boolean array may be enabled in any number of ways via the presence or absence of a chemical species describe a single bit, spatially patterning chemical's or physical markings to convey information, thresholding and binning the concentration of specific chemicals to describe some set of values in the Boolean array, or some combination of the given choices. The specifics of the physical encodings will be dictated by the needs of the specific good being marked and the desired cryptoledger scheme being implemented. For instance, if high throughput is needed to measure and mark bulk commodities like grain, food safe coatings of chemicals readable by infrared spectrometers could be added to grain as part of a simple but presumably limited authentication and cryptoledger scheme. Similar encoding methodology be used to mark petroleum products. Likewise for electronic goods a more sophisticated lithographic spatial and chemical patterning technique readable by a scanning electron microscope or similar instrument could be used to read the tags. At each stage of the supply chain the implanted array could be read and transferred (with the possibility of changing its modality) to new parts of the assembly as the manufacturing proceeds.

5.3 Future Work: Incorporating Digital Features into the Chemical Blockchain

Once a rudimentary chemo-physical ledger is established, it is possible that an end-user or their network might want to incorporate or integrate digital features with that same ledger. For instance, if one were using these cryptoledgers to enforce regulatory compliance through a supply chain, the various members of that supply chain might want to report their compliance to the government at each step of that chain. Alternatively, the various members of the supply chain might want to incorporate time stamps in the form of externally generated and recorded random numbers into the chemo-physical cryptoledger. In either of these cases, some form of digital communication is required to implement these ideas.

These desires present a twofold problem for this technology: the first problem is verifying and recording the external information being digitally broadcast, the second problem is transmitting information from the cryptoledger in a verifiable fashion. The solutions to both of these problems begins with the fact that physical objects are being handled. As such the goods being tracked by the chemo-physical cryptoledger are available for inspection and represent an inherent capital/reward for verification meaning that the transfer of the goods themselves as well as whatever reimbursement or rebate included as part of the execution of the supply chain are incentives to the various custodians along the way. While the specifications of the exact digital features themselves are currently unclear and the subject of future development, few, if any,

technical impediments are foreseen impacting the incorporation of digital features into a chemo-physical cryptoledger.

6. CONCLUSIONS

6.1 Further Research for Implementing the Chemical Blockchain

The past year of research has mainly focused upon developing the underlying theoretical technology for implementing a chemo-physical cryptoledger. Left out of this work has been the actual system specific implementations of underlying technology for physical implementations of chemo-physical cryptoledger. To transform this research effort into a practical working reality, much work needs to be directed at developing sensing and taggant technologies for specific classes of goods like petroleum products, commodity grains, electrical components, and pharmaceuticals. Specifically, tailored taggants need to be created for each of these classes of goods along with matched detection technologies that are appropriate for the specific mechanics of their logistics, regulatory requirements, and cost constraints

6.2 Possible End Users of the Chemical Blockchain

Earlier in this report, a variety of possible end users within the federal government for this technology have been proposed. With respect to tracking and validating supply chains, the DoD is has the greatest need for the security features of this technology, but other areas of government have possible current needs for this technology as well particularly when limited manpower constraints for enforcement are factored into the calculation. Among the agencies other than the DoD considered to be possible end users of a chemo-physical cryptoledger are: U.S. Customs and Border Protection (CBP), the Internal Revenue Service (IRS), the Environmental Protection Agency (EPA), the Food and Drug Administration (FDA), Alcohol Tobacco and Firearms (ATF), the Commodity Futures Trading Commission (CFTC), the U.S. State Department (DoS, the Transportation Security Administration (TSA), and the Department of Transportation (DoT).

Collectively, these departments and agencies have vast needs for verifying and validating all manner of goods for purposes of regulatory enforcement and security. These goods also

6.3 Possible Future Directions for the Chemical Blockchain

6.3.1 Extending Chemical Logic, Steganography, and Embedded Unimolecular Sensors to Applications in Government Cryptoledgers

The earlier discussion of chemical steganography [39], PUFs, chemical logic, and embedded unimolecular sensors raise the possibility of incorporating these technologies directly into the taggant-based cryptoledger system. These technologies could provide an independent, off-grid and difficult to detect chemical computer embedded in a variety of goods raising the possibility of significant intelligence gathering capabilities in settings that are typically very difficult to surveil. By having a supply chain capable of moving both goods and information in an otherwise mundane context, new and very difficult to disrupt vectors for information flow arise.

6.3.2 In Situ Chemical Sensing for Regulatory Enforcement

By adopting unimolecular sensing technology such as that associated with the chemical steganography described in the prior section as a vital working part of a chemo-physical ledger system, one might be able to incorporate forms of automatic *in situ* verification into a good itself so that the conditions of that good may be automatically recorded via encrypted hash into the cryptoledger. Implementing this idea would allow all manner of automatic enforcement which would present a bare minimum of a barrier for bad actors to have to overcome.

6.3.3 Interactions with Mechanism Design

Mechanism design [72–74] is a subfield of economics which inverts the usual precepts of game theory. Instead of asking what is the natural equilibrium/set of strategies associated with a set of rules, mechanism design seeks to make a given equilibrium or set of strategies the best what do the rules of the ”game” need to be. In developing and automatically enforcing various regulatory and compliance regimes, the interaction of these chemo-physical cryptoledger schemes potentially allows for complicated and targeted regulatory schemes that are currently cost prohibitive to enforce. By using chemo-physical cryptoledger technology to enforce and verify laws better, tailored governance could be implemented while maintaining privacy and enabling trust.

6.4 Final Remarks

In closing, this Karle Fellowship has enabled the theoretical background to a unique and valuable technology for government needs. The author looks forward to future work further developing this methodology in a practical setting and hopes that this work represents the beginning of ”bigger and better things” for this idea.

ACKNOWLEDGMENTS

Adam C. Knapp would like to gratefully acknowledge support via a Jerome and Isabella Karle Distinguished Scholar Fellowship via U.S. Naval Research Laboratory Section 219 funds.

REFERENCES

1. D. . A. A. Office of the Assistant Secretary of the Navy (Research and B. Management, *COUNTERFEIT MATERIEL PROCESS GUIDEBOOK: Guidelines for Mitigating the Risk of Counterfeit Materiel in the Supply Chain* (U.S. Government: Department of Navy, 2017).

2. S. Nakamoto et al., “Bitcoin: A peer-to-peer electronic cash system (2008).

3. xxx, “Bitcoin blockchain - date of genesis block’s mining,” `https://bitcoin.org/en/download`” (cited July 2019).

4. L. Kugler, “Why cryptocurrencies use so much energy: and what to do about it,” *Communications of the ACM* **61**(7), 15–17 (2018).

5. S. Elnaj, "The problems with bitcoin and the future of blockchain," https://www.forbes.com/sites/forbestechcouncil/2018/03/29/the-problems-with-bitcoin-and-the-future-of-blockchain/#5631581668dc" (cited August 26, 2019).

6. Anonymous, "Bitcoin problems 2019 - and what is being done," https://cryptalker.com/bitcoin-problems/" (cited August 26, 2019).

7. Z. Li and Q. Liao, "Toward socially optimal bitcoin mining," in *2018 5th International Conference on Information Science and Control Engineering (ICISCE)*, pp. 582–586 (IEEE, 2018).

8. P. Chatzigiannis, F. Baldimtsi, I. Griva, and J. Li, "Diversification across mining pools: Optimal mining strategies under pow," *arXiv preprint arXiv:1905.04624* (2019).

9. J. Wilmoth, "Bitmain's mining pools now control nearly 51 percent of the bitcoin hashrate," https://www.ccn.com/bitmains-mining-pools-now-control-nearly-51-percent-of-the-bitcoin-hashrate/" (cited August 27, 2019).

10. W. Zhao, "It's now harder to mine bitcoin than ever," https://www.coindesk.com/bitcoin-hash-rate-new-record" (cited August 27, 2019).

11. J. M. Connor, "Global price fixing: Second paperback edition" (2008).

12. M. Frankel, "The 7 biggest challenges facing bitcoin," https://www.fool.com/investing/2017/11/19/the-7-biggest-challenges-facing-bitcoin.aspx" (cited August 26, 2019).

13. T. M. Fernández-Caramés and P. Fraga-Lamas, "A review on the use of blockchain for the internet of things," *IEEE Access* **6**, 32979–33001 (2018).

14. D. Puthal and S. P. Mohanty, "Proof of authentication: Iot-friendly blockchains," *IEEE Potentials* **38**(1), 26–29 (2018).

15. V. Saini, "Consensuspedia: An encyclopedia of 30+ consensus algorithms," https://hackernoon.com/consensuspedia-an-encyclopedia-of-29-consensus-algorithms-e9c4b4b7d08f" (cited August 27, 2019).

16. E. W. Community, "Ethereum wiki - problems," https://github.com/ethereum/wiki/wiki/Problems" (cited August 27, 2019).

17. D. S. L. . UIUC, ""fake stake" attacks on chain-based proof-of-stake cryptocurrencies," https://medium.com/@dsl_uiuc/fake-stake-attacks-on-chain-based-proof-of-stake-cryptocurrencies-b8b05723f806" (cited August 27, 2019).

18. Z. Zhang, V. Vasavada, R. King, and L. Zhang, "Proof of authentication for private distributed ledger," in *Proceedings of the NDSS Workshop on Decentralised IoT Systems and Security (DISS)* (2019).

19. J. Gooch, B. Daniel, V. Abbate, and N. Frascione, "Taggant materials in forensic science: A review," *TrAC Trends in Analytical Chemistry* **83**, 49–54 (2016).

20. C. Smith, S. Strauss, and L. DeFrancesco, "Dna goes to court," *Nature biotechnology* **30**(11), 1047 (2012).

21. D. Starr, "Forensics gone wrong: When dna snares the innocent," *Science* (Mar 2016), doi: 10.1126/science.aaf4160, URL http://dx.doi.org/10.1126/science.aaf4160.

22. B. Duong, H. Liu, C. Li, W. Deng, L. Ma, and M. Su, "Printed multilayer microtaggants with phase change nanoparticles for enhanced labeling security," *ACS applied materials & interfaces* **6**(11), 8909–8912 (2014).

23. U. F. Government, "U.s. currency education program," https://www.uscurrency.gov/denominations/100" (cited August 20, 2019").

24. S.-G. of the International Civil Aviation Organization, "Convention on the marking of plastic explosives for the purpose of detection," https://treaties.un.org/doc/db/Terrorism/Conv10-english.pdf" (cited August 20, 2019).

25. N. Kaish, J. Fraser, V. Otugen, and S. Popovic, "Method for remote detection of volatile taggant" (Feb. 15 2000), US Patent 6,025,200.

26. SmartWater, "Smartwater," https://www.smartwater.com" (cited August 20, 2019).

27. SelectaDNA, "Selectadna": Advanced forensic marking," https://www.selectadna.co.uk/atm-gas-attacks/selectadna-surge" (cited August 20, 2019).

28. B. Schneier, "Schneier on security: Smart water," https://www.schneier.com/blog/archives/2005/02/smart_water.html" (cited August 22, 2019).

29. I. MicroTrace Solutions, "Microtrace solutions, inc.," https://www.microtracesolutions.com/taggant-technologies" (cited August 22, 2019).

30. R. Committee on Marking, M. Commission on Physical Sciences, D. Sciences, and N. Council, *Marking, Rendering Inert, and Licensing of Explosive Materials: Interim Report* (National Academies Press, 1997), ISBN 9780309590587, URL https://books.google.com/books?id=5v4rXnCNxzcC.

31. A. Glover, N. Aziz, J. Pillmoor, D. W. McCallien, and V. B. Croud, "Evaluation of dna as a taggant for fuels," *Fuel* **90**(6), 2142–2146 (2011).

32. E. Bellaver, "Anti-counterfeiting technology taggants and explosives: How high a priority is anti-counterfeiting?," https://blog.thelabelprinters.com/84" (cited September 24, 2019).

33. C. Goldsberry, "Material taggants provide protection from counterfeiting of plastic products," https://www.plasticstoday.com/injection-molding/material-taggants-provide-protection-counterfeiting-plastic-products/30146431923299" (cited September 24, 2019).

34. M. KGaA, "Securalic - the innovative way to safeguard brand property," https://www.emdgroup.com/en/brands/pm/securalic.html" (cited September 24, 2019).

35. A. D. Sciences, "Applied dna sciences," https://adnas.com/" (cited September 24, 2019).

36. J. Andréasson and U. Pischel, "Molecules for security measures: from keypad locks to advanced communication protocols," *Chemical Society Reviews* **47**(7), 2266–2279 (2018).

37. D. Margulies, C. E. Felder, G. Melman, and A. Shanzer, "A molecular keypad lock: a photochemical device capable of authorizing password entries," *Journal of the American Chemical Society* **129**(2), 347–354 (2007).

38. J. Andréasson, S. D. Straight, T. A. Moore, A. L. Moore, and D. Gust, "An all-photonic molecular keypad lock," *Chemistry–A European Journal* **15**(16), 3936–3939 (2009).

39. T. Sarkar, K. Selvakumar, L. Motiei, and D. Margulies, "Message in a molecule," *Nature communications* **7**, 11374 (2016).

40. O. Guillou, C. Daiguebonne, G. Calvez, and K. Bernot, "A long journey in lanthanide chemistry: from fundamental crystallogenesis studies to commercial anticounterfeiting taggants," *Accounts of chemical research* **49**(5), 844–856 (2016).

41. J. F. Stoddart, X. Hou, and C. Ke, "Supramolecular fluorescent dyes" (May 19 2016), US Patent App. 14/867,826.

42. C. P. Carvalho, Z. Domínguez, J. P. Da Silva, and U. Pischel, "A supramolecular keypad lock," *Chemical Communications* **51**(13), 2698–2701 (2015).

43. U. Rührmair, J. Sölter, and F. Sehnke, "On the foundations of physical unclonable functions.," *IACR Cryptology ePrint Archive* **2009**, 277 (2009).

44. M. Naor and A. Shamir, "Visual cryptography," in *Workshop on the Theory and Application of of Cryptographic Techniques*, pp. 1–12 (Springer, 1994).

45. E. R. Verheul and H. C. Van Tilborg, "Constructions and properties of k out of n visual secret sharing schemes," *Designs, Codes and Cryptography* **11**(2), 179–196 (1997).

46. G. Ateniese, C. Blundo, A. De Santis, and D. R. Stinson, "Extended capabilities for visual cryptography," *Theoretical Computer Science* **250**(1-2), 143–161 (2001).

47. M. Gnanaguruparan and S. Kak, "Recursive hiding of secrets in visual cryptography," *Cryptologia* **26**(1), 68–76 (2002).

48. R. Ito, H. Kuwakado, and H. Tanaka, "Image size invariant visual cryptography," *IEICE transactions on fundamentals of electronics, communications and computer sciences* **82**(10), 2172–2177 (1999).

49. J. Weir and W. Yan, *A Comprehensive Study of Visual Cryptography*, pp. 70–105 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010), ISBN 978-3-642-14298-7, doi: 10.1007/978-3-642-14298-7_5, URL `https://doi.org/10.1007/978-3-642-14298-7_5`.

50. Y.-C. Hou, "Visual cryptography for color images," *Pattern recognition* **36**(7), 1619–1629 (2003).

51. R. De Prisco and A. De Santis, "Color visual cryptography schemes for black and white secret images," *Theoretical Computer Science* **510**, 62–86 (2013).

52. D. J. Wu, "Fully homomorphic encryption: Cryptography's holy grail," *XRDS* **21**(3), 24–29 (Mar. 2015), ISSN 1528-4972, doi: 10.1145/2730906, URL `http://doi.acm.org/10.1145/2730906`.

53. X. Liu, S. Wang, X. Yan, and W. Zhang, "Homomorphic visual cryptography," *Journal of Information Hiding and Multimedia Signal Processing* **8**, 744–756 (01 2017).

54. X. Yan, Y. Lu, L. Liu, S. Wan, W. Ding, and H. Liu, "Exploiting the homomorphic property of visual cryptography," *International Journal of Digital Crime and Forensics (IJDCF)* **9**(2), 45–56 (2017).

55. X. Yan, Y. Lu, and L. Liu, "A general progressive secret image sharing construction method," *Signal Processing: Image Communication* **71**, 66–75 (2019).

56. R. L. Rivest, L. Adleman, M. L. Dertouzos, et al., "On data banks and privacy homomorphisms," *Foundations of secure computation* **4**(11), 169–180 (1978).

57. C. Gentry et al., "Fully homomorphic encryption using ideal lattices.," in *Stoc*, volume 9, pp. 169–178 (2009).

58. C. Gentry, *A fully homomorphic encryption scheme*, PhD thesis (Stanford University, 2009), `crypto.stanford.edu/craig`.

59. B. Hayes, "Alice and bob in cipherspace," *American Scientist* **100**(5), 362–367 (2012).

60. L. Lamport, "The part-time parliament," *ACM Transactions on Computer Systems (TOCS)* **16**(2), 133–169 (1998).

61. L. Lamport et al., "Paxos made simple," *ACM Sigact News* **32**(4), 18–25 (2001).

62. A. Kancherla, "Paxos made simple. for real," `https://medium.com/@nevverlander/paxos-made-simple-for-real-aa221be7d91b`" (cited September 19, 2019).

63. C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in S. Halevi and T. Rabin, eds., *Theory of Cryptography*, pp. 265–284, Berlin, Heidelberg, 2006 (Springer Berlin Heidelberg), ISBN 978-3-540-32732-5.

64. C. Dwork, "Differential privacy," in M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., *Automata, Languages and Programming*, pp. 1–12, Berlin, Heidelberg, 2006 (Springer Berlin Heidelberg), ISBN 978-3-540-35908-1.

65. C. Voemel, F. De Lorenzi, S. Beer, and E. Fuchs, "The secret life of keys: on the calculation of mechanical lock systems," *SIAM Review* **59**(2), 393–422 (2017).

66. Z. Zhang, V. Vasavada, X. Ma, and L. Zhang, "Dledger: An iot-friendly private distributed ledger system based on dag," *arXiv preprint arXiv:1902.09031* (2019).

67. C. Fan, *Performance Analysis and Design of an IoT-Friendly DAG-based Distributed Ledger System*, PhD thesis (University of Alberta, 2019).

68. P. D'Arco and R. De Prisco, "Secure computation without computers," *Theoretical Computer Science* **651**, 11–36 (2016).

69. A. Churyumov, "Byteball: A decentralized system for storage and transfer of value," *URL https://byteball. org/Byteball. pdf* (2016).

70. C. LeMahieu, "Nano: A feeless distributed cryptocurrency network," *URL: https://nano. org/en/whitepaper* (2018).

71. S. Popov, "The tangle," *cit. on* p. 131 (2016).

72. F. McSherry and K. Talwar, "Mechanism design via differential privacy.," in *FOCS*, volume 7, pp. 94–103 (2007).

73. T. Börgers, *An introduction to the theory of mechanism design* (Oxford University Press, USA, 2015).

74. R. V. Vohra, *Mechanism design: a linear programming approach*, volume 47 (Cambridge University Press, 2011).

75. F. Nielsen and R. Nock, "Clustering multivariate normal distributions," in *Emerging Trends in Visual Computing*, pp. 164–174 (Springer, 2009).

76. P. Olver, *Applications of Lie Groups to Differential Equations*, Applications of Lie Groups to Differential Equations (Springer New York, 2000), ISBN 9780387950006, URL https://books.google.com/books?id=sI2bAxgLMXYC.

77. L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions* (Springer Netherlands, 2012), ISBN 9789401021968, URL https://books.google.com/books?id=PJzuCAAAQBAJ.

78. G. Arfken, H. Weber, and F. Harris, *Mathematical Methods for Physicists: A Comprehensive Guide* (Elsevier Science, 2013), ISBN 9780123846549.

79. F. Oberhettinger, *Tables of Mellin Transforms* (Springer Berlin Heidelberg, 2012), ISBN 9783642659751.

80. H. Buchholz, H. Lichtblau, and K. Wetzel, *The Confluent Hypergeometric Function: with Special Emphasis on its Applications*, Springer Tracts in Natural Philosophy (Springer Berlin Heidelberg, 2013), ISBN 9783642883965.

81. L. U. Ancarani and G. Gasaneo, "Derivatives of any order of the hypergeometric function p f q (a 1 , ..., a p ; b 1 , ..., b q ; z) with respect to the parameters a i and b i," *Journal of Physics A: Mathematical and Theoretical* **43**(8), 085210 (2010), URL http://stacks.iop.org/1751-8121/43/i=8/a=085210.

82. A. ZUCCOLO, "Wronskian representations of hypergeometric integrals (2018).

83. E. Jones, T. Oliphant, P. Peterson, et al., "SciPy: Open source scientific tools for Python" (2001–), URL http://www.scipy.org/, [Online; accessed Sept. 21, 2018].

Appendix A

JEROME AND ISABELLA KARLE DISTINGUISHED SCHOLAR FELLOWSHIP PROPOSAL STATEMENT

Development of a Chemical Blockchain with Biased Detection of Matched Taggants

Adam C. Knapp, Code 6181

OBJECTIVE: The objective of this work is to explore and develop an approach to matched chemical taggant/detection systems to create a chemically-encrypted physical, blockchain scheme for Navy and DoD-relevant problems in trusted supply chain management and regulatory enforcement.

BACKGROUND: The need for control over complicated equipment and supply chains is critical for the U.S. Navy, as it often involves international and/or potentially untrusted partners. The use of a blockchain, (an encrypted, updateable, and trustless database) has been proposed as a digital solution for this problem. Unfortunately, this approach often imposes a high computational/energetic cost making it unsuitable for small or moderately sized production runs or situations where energy efficiency/low radiation is important. It also lacks an intrinsic, physical component for multifactor authentication and verification. Likewise, physical taggant technologies developed by the private sector for anti-counterfeiting and theft deterrence purposes often fail to associate the digital with the physical, leaving the underlying databases open to independent physical or digital attack.

TECHNICAL APPROACH: This work will focus on adapting the theory underlying the digital blockchain and suitable cryptographic systems to physical, chemical taggant/detector systems. To focus the effort, small, high-value goods like microelectronics will serve as a hypothetical test case. The program, itself, will center on developing theoretical techniques to enable goods to be marked with chemical taggants encoding an encrypted database during manufacture and able to be updated throughout the various steps in the supply chain. The relationship between candidate taggant systems and blockchain capability will then be modeled. These taggants will serve as a physical platform for multifactor authentication and relevant encrypted database storage. The requirements for accurate and robust detection/readout of the chemical blockchain will be theoretically determined using information theoretic models. This effort will inform the ultimate feasibility of the approach for specific applications since more complex taggant systems may require more robust and costly chemical detection methods. Next, various approaches for physically updating the taggant signature as it moves through a hypothetical supply chain will be investigated and developed. Lastly, methodology for digitally and/or physically decrypting the taggant encoded databases in a selective and hierarchically controlled, need-to-know fashion will be developed.

PLANNED ACCOMPLISHMENTS AND MANNER IN WHICH THEY WILL BE MEASURED: Cryptographic techniques like visual cryptography will be generalized and transferred from a digital domain to physio-chemical one to enable a blockchain-like hierarchically controlled, encrypted, updateable database scheme. Chemical taggant systems will be identified and matched to appropriate biased detection systems. Theoretical understanding and metrics will be developed, enabling clear assessment of capabilities and limitations of the proposed technology for practical implementation. Results will be disseminated via technical reports, peer-reviewed journals, and conferences.

BENEFITS TO SCIENCE AND TECHNOLOGY: If successful, this technology could be immediately applicable to goods such as semiconductors and microelectronics, which currently suffer from serious counterfeiting and supply-chain hygiene problems. This approach could also address regulatory enforcement problems outside of the DoD, such as customs and taxation. This project will significantly assist the U.S. Navy in managing, securing, and validating technology for logistics and supply chains.

Appendix B

9TH ANNUAL IRS-TPC JOINT RESEARCH CONFERENCE ON TAX ADMINISTRATION ABSTRACT APPLICATION: SENT DECEMBER 3, 2018

++ A Physically Decentralized Ledger Scheme for Tax and Regulatory Compliance and Enforcement

Adam C. Knapp, Ph. D.
E-mail: adam.knapp@nrl.navy.mil; Tel: +1 (202) 404 5487
U. S. Naval Research Laboratory, Chemistry Division, Chemical Sensing and Fuel Technology, Code 6181
4555 Overlook Ave SW, Washington, DC 20375 USA

Government agencies, like the IRS, are being asked to do more with less in increasingly complex operating environments. Frequently, these agencies have complicated missions taking place in extended social and business networks subject to overlapping jurisdictions and regulations. Unsurprisingly, given this complexity, resource constraints and heterogeneous operating environments often force the government to make difficult enforcement decisions. Within the context of taxation, these challenges lead to some taxes, either individually or categorically, not being collected, which encourages games of chicken between taxpayers and tax collectors and can leave compliant taxpayers to view the system as rigged or inherently unfair.

Directly addressing these problems by providing tax (and other regulatory) authorities with a scheme which encourages and secures compliance while reporting granular-level knowledge of taxable/regulatory events is the ultimate goal of this work. The remainder of this summary outlines the key mechanics and features of this project. We begin by first describing, the problems addressed by a hypothetical secured and decentralized ledger as well as sketch an outline of how it might be used. A high-level technical description of the implementation of such a secure ledger then follows.

Tax authorities have long known that full compliance in self-reporting tax regimes is highly correlated to the presence of income reporting and withholding schemes. In effect, external auditing and payments systems work as much higher compliance is found when third parties like employers and banks act as reporting/withholding agents. Self-reporting is used because generally tax authorities do not have access to externally verifiable transaction data since many, if not most, taxable events are not as predictable as payroll. In an audit, unrecorded or inaccessible financial records may complicate enforcement efforts by limiting the availability of relevant information.

VAT regimes address these problems by having a flat tax rate applied to the value-added portion of a product or service which is collected and credited incrementally throughout a supply chain. Ultimately, only the end-consumer pays the tax, but does not receive the credit. While VAT methods effectively leverage a supply chains network to report and collect on taxable events, they are unable to easily handle changing jurisdictions or account for varying tax rates and credits. If they could, missing trader and carousel fraud schemes would not be as problematic as they are to VAT collecting countries. Associating an enciphered, updateable ledger comprised of physical and digital components with trade goods at either the individual or lot level would allow a tax authority as well as taxpayer(s) to track and verify tax and business data throughout a supply chain. This decentralized ledger could be hierarchically controlled and updated both

physically and digitally, and thus, provide dual digital and physical means of protecting against malfeasance as well as communicate a more richly detailed level of transaction data. Tampering with such a system would be not just a criminal act against the state, but a civil wrong as well since noncompliance would be a form of fraud perpetrated against business partners. Tax authorities would thus be able to indirectly use the civil courts to enforce tax collection.

While the implementation of the digital portion of such a ledger could be done with standard cryptographic tools, the physical portion of such a ledger requires a variety of tagging and tagging detection methods. These tagging systems like RFID chips/scanners, additive manufacturing tags/IR detectors, or liquid taggants/uv-vis detectors, would depend upon the specific type of trade good in question. Physical cryptographic techniques like visual cryptography could be adapted for taxation and provide a known platform for supporting the updating of the ledgers without decipherment (homomorphic encryption). These physical tags would also be the perfect platform to use random physical processes like percolation or polymer cracking to create tags that would be nearly impossible to counterfeit. They could be initiated using starting configurations publicly posted by tax authorities at known times which would provide a trustworthy temporal physical record of the supply chain and its value addition.

Ultimately, given the state of current taggant technology, low-cost embedded mixed physical-digital ledger systems could be created that enable fiscally and manpower efficient tax collection. They would enable tax collection with a known, enforceable, and actionable audit trail which arises naturally as the byproduct of the business process with each member of that business network empowered and financially incentivized to make sure taxes are properly, correctly, and completely collected.

Appendix C

DERIVATIONS ASSOCIATED FOR AN UNUSED ALTERNATIVE NOISE AND INTEFERENT MODEL FOR CRYPTOLEDGER CHEMICAL DETECTION

C.1 Author's Note

Not all research directions work. When I first started work on this Karle Fellowship-funded project, I thought that there would be a significant amount of work to be done associated with things like noise detection models and chemical interferents. I had initially thought that I would essentially be modifying an existing internet of things framework to a chemical detection setting and that most of the work would detail how to handle things like chemical interferents and problem specific detection settings. However, as I dug deeper into the chemical inspection, computer science, and electrical engineering literature, I discovered that many of the pieces I had thought existed for the internet of things were still unsettled research areas and that while the pieces I wanted existed, they needed to be adapted and reconfigured for a chemical detection setting. Moreover, since the chemical and physical nature of the chemo-physical cryptoledger is presumably a localized and engineered environment it can naturally adapted and tailored to specific detection technologies, so the research direction represented by concerns about chemical interferents and noise models is presumably less necessary at a basic research level than I had initially anticipated. Nonetheless, some of this work was performed as I attempted to "hedge my bets" as the situation became clearer to me. This work is included to document its existence in support of the Karle Fellowship directive's documentation requirements. The calculations included below are intended to support the calculation of the Kullback-Leibler divergence between for linear response sensor arrays in the presence of chemical interferents with concentrations present according to a shifted Gaussian functions as a means to measure the similarity or quality of a sensor array.

C.2 Quality of a Chemical Simulant for a Linear Response Sensor Array with Gaussian Noise Using Synthetic Data

For this section, two multivariate normal distributions, $\mathcal{N}_0$ and $\mathcal{N}_1$, where

$$\mathcal{N}_i(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = |\pi\boldsymbol{\Sigma}_i|^{-1/2} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)} \tag{C1}$$

are the probability distribution input into the KLD. The resultant expression is given in closed form [75] as

$$\begin{aligned} D_{KL}(\mathcal{N}_0||\mathcal{N}_1) &= \int_{-\infty}^{\infty} d\boldsymbol{x} \mathcal{N}_0(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \ln\left(\frac{\mathcal{N}_0(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)}{\mathcal{N}_1(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)}\right) \\ &= \frac{1}{2}\left(\mathrm{Tr}(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) - M + \ln\left(\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|}\right)\right) \end{aligned} \tag{C2}$$

where M is the number of sensors in the sensor array, $\mathcal{N}_0$ is the probability distribution of sensor responses of the sensor array when the target analyte is present and not the simulant, $\mathcal{N}_1$ is the probability distribution

of sensor responses of the sensor array when the CS is present, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the mean vector and covariance matrix of their respective Gaussian distribution.

If one supposes an environment without the presence of chemical interferents, then using the KLD to measure the quality of a CS-target analyte system using a sensor array with noise well-described by a multivariate normal distribution model is relatively simple as is optimizing the design of a CS by minimizing the KLD.

C.3 Mathematical Description of the Models Used for the Environmental Unknowns/Chemical Noise

The effects of an unknown chemical background environment upon our chemical simulant and detection system may be modeled as follows

$$P_{S+E}(\boldsymbol{s},\boldsymbol{c},\boldsymbol{x};\boldsymbol{a},\boldsymbol{\sigma},\boldsymbol{p},\boldsymbol{q}) = P_S(\boldsymbol{s};\boldsymbol{a},\boldsymbol{\sigma},\boldsymbol{c}_0|\boldsymbol{c},\boldsymbol{x})P_E(\boldsymbol{c},\boldsymbol{x};\boldsymbol{p},\boldsymbol{q}) \tag{C3}$$

where $P_E(\boldsymbol{c},\boldsymbol{x};\boldsymbol{p},\boldsymbol{q})$ is the probability distribution associated with a particular chemical background environment and $P_S(\boldsymbol{s};\boldsymbol{a},\boldsymbol{\sigma},\boldsymbol{c}_0|\boldsymbol{c},\boldsymbol{x})$ is the probability distribution associated with the response of the sensing system and conditioned upon the particular environment defined by P_E. $\boldsymbol{p}$ and $\boldsymbol{q}$ are parameters which define the chemical background environment and which may be conditioned on average environmental concentrations and their variances of various chemical interferents. $\boldsymbol{x}$

$$\begin{aligned} A_1 &= \int_0^\infty dc\, c e^{-\frac{(c-p)^2}{2q^2}} \\ &= \left[\sqrt{\frac{\pi}{2}} pq \operatorname{erf}\left(\frac{c-p}{\sqrt{2}q}\right) - q^2 e^{-\frac{(c-p)^2}{2q^2}}\right]_0^\infty \\ &= \sqrt{\frac{\pi}{2}} pq\left(1+\operatorname{erf}\left(\frac{p}{\sqrt{2}q}\right)\right) - q^2 e^{-\frac{p^2}{2q^2}} \end{aligned} \tag{C4}$$

$$\begin{aligned} A_2 =& A_1\langle x\rangle_1 = \int_0^\infty dc\, c^2 e^{-\frac{(c-p)^2}{2q^2}} \\ &= \left[\sqrt{\frac{\pi}{2}} q(p^2+q^2)\operatorname{erf}\left(\frac{c-p}{\sqrt{2}q}\right)\right]_0^\infty \\ &\quad - \left[q^2 e^{-\frac{(c-p)^2}{2q^2}}\right]_0^\infty \\ &= \sqrt{\frac{\pi}{2}} p(p^2+q^2)\left(1+\operatorname{erf}\left(\frac{p}{\sqrt{2}q}\right)\right) + pq^2 e^{-\frac{p^2}{2q^2}} \end{aligned} \tag{C5}$$

$$\begin{aligned}
A_2\langle x\rangle_2 =& A_1\langle x^2\rangle_1 = \int_0^\infty dc\, c^3 e^{-\frac{(c-p)^2}{2q^2}} \\
&= \left[\sqrt{\frac{\pi}{2}} pq(p^2+3q^2)\,\mathrm{erf}\left(\frac{c-p}{\sqrt{2}q}\right)\right]_0^\infty \\
&\quad - \left[q^2 e^{-\frac{(c-p)^2}{2q^2}}(p^2+pc+2q^2+c^2)\right]_0^\infty \\
&= \sqrt{\frac{\pi}{2}} pq(p^2+3q^2)\left(1+\mathrm{erf}\left(\frac{p}{\sqrt{2}q}\right)\right) \\
&\quad + q^2 e^{-\frac{p^2}{2q^2}}(p^2+2q^2)
\end{aligned} \tag{C6}$$

$$\begin{aligned}
A_2\langle x^2\rangle_2 &= \int_0^\infty dc\, c^4 e^{-\frac{(c-p)^2}{2q^2}} \\
&= \left[\sqrt{\frac{\pi}{2}} q(p^4+6p^2q^2+3q^4)\,\mathrm{erf}\left(\frac{c-p}{\sqrt{2}q}\right)\right]_0^\infty \\
&\quad - \left[q^2 e^{-\frac{(c-p)^2}{2q^2}}(p^3+p^3c+2p(5q^2+c^2)+3q^2c+c^3)\right]_0^\infty \\
&= \sqrt{\frac{\pi}{2}} q(p^4+6p^2q^2+3q^4)\left(1+\mathrm{erf}\left(\frac{p}{\sqrt{2}q}\right)\right) \\
&\quad + q^2 e^{-\frac{p^2}{2q^2}}(p^3+5pq^2)
\end{aligned} \tag{C7}$$

Combining eqns. (**??**,**??**, C4, C5, C6, C7, C39), allow us to calculate the normalization of eqn. (**??**) as well as set the parameters p_j and q_j in terms of the variance and mean concentration of the interferent when it is present in the environment.

Due to the complicated expressions associated with variance and mean concentration, parameters p_j and q_j must be solved for numerically.

The probability associated with the presence or absence of chemical interferent is calculated as follows

$$\sum_{\boldsymbol{x}\in[\mathbf{0},\mathbf{1}]^M} p(\boldsymbol{x}) = \sum_{\boldsymbol{x}\in[\mathbf{0},\mathbf{1}]^M}\left(\prod_{j=1}^M p(x_j)\right) \tag{C8}$$

$$\prod_{j=1}^M\left(\int_0^\infty dc_j\, c_j e^{-\frac{(c_j-p_j)^2}{2q_j^2}}\right) = \tag{C9}$$

$$\prod_{j=1}^{M}\left(\int_0^{\infty} dc_j\, c_j^2 e^{-\frac{(c_j-p_j)^2}{2q_j^2}}\right) = \tag{C10}$$

$$\begin{aligned}&\left(\sum_{\boldsymbol{x}\in[\mathbf{0},\mathbf{1}]^N} p(\boldsymbol{x})\prod_{i=1}^{N}\int_{-\infty}^{\infty} ds_i\, e^{-\frac{(s_i-\mu_i)^2}{2\sigma_i^2}}\right)\\ &= (2\pi)^{N/2}\left(\sum_{\boldsymbol{x}\in[\mathbf{0},\mathbf{1}]^N} p(\boldsymbol{x})\sqrt{\prod_{i=1}^{N}\sigma_i^2}\right)\end{aligned} \tag{C11}$$

Recalling eqns. (**??**) and (**??**) which are repeated here as:

$$P_{D+E}\left(\boldsymbol{X}_{D+E};\boldsymbol{\theta}_{D+E}\right) = P_D\left(\boldsymbol{X}_D|\boldsymbol{X}_E;\boldsymbol{\theta}_D\right)P_E\left(\boldsymbol{X}_E;\boldsymbol{\theta}_E\right) \tag{C12}$$

where

$$P_E\left(\boldsymbol{X}_E;\boldsymbol{\theta}_E\right) = P_{\boldsymbol{c}}\left(\boldsymbol{c}|\boldsymbol{a};\boldsymbol{A}\right)P_{\boldsymbol{a}}\left(\boldsymbol{a};\boldsymbol{A}\right) \tag{C13}$$

Focusing first on $P_{\boldsymbol{a}}\left(\boldsymbol{a};\boldsymbol{A}\right)$, we model the presence or absence of a background by boolean vector $\boldsymbol{a}$ describing generated via a Poisson sampling process. In a Poisson process each

C.4 Indefinite Gaussian Integrals

Calculations for both the Kullback-Leibler divergence as well as the normalizations associated with the environmental background probability distributions are reliant upon the indefinite integration of an arbitrary Gaussian moment. Consequently, we introduce this appendix by first calculating these moments for subsequent reference.

First, we present the well-known indefinite Gaussian integral,

$$\int dx\, e^{-ax^2+bx} = \frac{1}{2}\sqrt{\frac{\pi}{a}}\, e^{\frac{b^2}{4a}}\operatorname{erf}\left(\frac{2ax-b}{2\sqrt{a}}\right) \tag{C14}$$

where a and b are arbitrary parameters. To calculate the moments of this indefinite integral, we differentiate with respect to b to obtain the following,

$$\begin{aligned}\int dx\, x^n e^{-ax^2+bx} &= \frac{\partial^n}{\partial b^n}\int dx\, e^{-ax^2+bx}\\ &= \frac{1}{2}\sqrt{\frac{\pi}{a}}\frac{\partial^n}{\partial b^n}\left[e^{\frac{b^2}{4a}}\operatorname{erf}\left(\frac{2ax-b}{2\sqrt{a}}\right)\right]\end{aligned} \tag{C15}$$

which may then be rewritten using the general Leibniz formula, $(fg)^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)}(x) g^{(k)}(x)$ [76], to give,

$$\int dx x^n e^{-ax^2+bx} = \frac{1}{2}\sqrt{\frac{\pi}{a}} \sum_{k=0}^{n} \binom{n}{k} \frac{\partial^{n-k}}{\partial b^{n-k}} \left[e^{\frac{b^2}{4a}} \right] \cdot \frac{\partial^k}{\partial b^k} \left[\text{erf}\left(\frac{2ax-b}{2\sqrt{a}} \right) \right] \tag{C16}$$

To calculate the remaining derivative terms, we first use Faà di Bruno's formula [77], to express the derivative of the Gaussian term as

$$\frac{\partial^n}{\partial b^n} \left[e^{\frac{b^2}{4a}} \right] = e^{\frac{b^2}{4a}} \sum_{j=0}^{n} B_{n,j} \left(\frac{b}{2a}, \frac{1}{2a}, 0, \ldots, 0 \right) \tag{C17}$$

where

$$\begin{aligned} B_{n,k}(x_1, \ldots, x_{n-k+1}) = n! \sum_{j} \left(\prod_{i=1}^{n-k+1} \left(\frac{1}{j_i!} \left(\frac{x_i}{i!} \right)^{j_i} \right) \right) \\ \text{s.t. } \sum_i j_i = k \text{ and } \sum_i i j_i = n \end{aligned} \tag{C18}$$

are the Bell polynomials [77]. The derivatives of the error function term may be found with the help of Rodrigues' formula,

$$\frac{\partial^n}{\partial b^n} \text{erf}\left(\frac{2ax-b}{2\sqrt{a}} \right) = \begin{cases} \text{erf}\left(\frac{2ax-b}{2\sqrt{a}} \right) & \text{if } n = 0 \\ \frac{-2}{(2\sqrt{a})^n \sqrt{\pi}} H_{n-1}\left(\frac{2ax-b}{2\sqrt{a}} \right) e^{-\frac{(2ax-b)^2}{4a}} & \text{if } n \geq 1 \end{cases} \tag{C19}$$

where

$$H_n(x) = n! \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^j}{j!(n-2j)!} (2x)^{n-2j} \tag{C20}$$

are the so-called physicists' Hermite polynomials [78].

C.5 Derivation of the Kullback-Leibler Divergence for Gaussian Sensors in a Noisy Background

Evaluating the KLD either analytically or numerically is the key component in calculating the proposed CS FOM. Since the proposed probabilistic model is comprised of discrete and continuous variables, we consider the KLD formulation for mixed discrete and continuous probability distributions given by,

$$D(P_0(\boldsymbol{x}, \boldsymbol{y}) || P_1(\boldsymbol{x}, \boldsymbol{y})) = \sum_{\boldsymbol{x}} \int d\boldsymbol{y} P_0(\boldsymbol{x}, \boldsymbol{y}) \ln \left(\frac{P_0(\boldsymbol{x}, \boldsymbol{y})}{P_1(\boldsymbol{x}, \boldsymbol{y})} \right) \tag{C21}$$

The joint probability distribution for this paper's model for a chemical sensing system and its surrounding environment is given by

$$P_{S+E_n}(\boldsymbol{c},\boldsymbol{x};\boldsymbol{c}_0,\boldsymbol{a},\boldsymbol{\sigma},\boldsymbol{p},\boldsymbol{q},\boldsymbol{n}) = \left(\frac{p(\boldsymbol{x})}{\mathcal{N}_{S+E_n}}\right)\cdot\left(\prod_{i=1}^{N} e^{-\frac{(s_i-\mu_i(\boldsymbol{c},\boldsymbol{x};\boldsymbol{c}_0,\boldsymbol{a}))^2}{2\boldsymbol{v}(\boldsymbol{x};\boldsymbol{\sigma})_i^2}}\right)\cdot\left(\prod_{j=1}^{M} c_j^{n_j} e^{-\frac{(c_j-p_j)^2}{2q_j^2}}\right) \tag{C22}$$

where $p(\boldsymbol{x})$ is the probability distribution of binary variables signifying the presence or absence of various chemical interferents, $\prod_{j=1}^{M} c_j^{n_j} e^{-\frac{(c_j-p_j)^2}{2q_j^2}}$ denotes the probability distribution for the concentrations of the background chemical interferents, and $\prod_{i=1}^{N} e^{-\frac{(s_i-\mu_i(\boldsymbol{c},\boldsymbol{x};\boldsymbol{c}_0,\boldsymbol{a}))^2}{2\boldsymbol{v}(\boldsymbol{x};\boldsymbol{\sigma})_i^2}}$ is the probability of the sensor response given specific environmental factors.

Next, we consider the logarithmic portion of the KLD given by eqn. (C22) as both of the probability distributions associated with the target analyte and the simulant conform to the model given by eqn. (C22),

$$\begin{aligned}\langle \ln P_{S+E_n}\rangle =& \langle \ln p(\boldsymbol{x})\rangle_{S+E_n} - \langle \ln \mathcal{N}_{S+E_n}\rangle_{S+E_n} \\ &- \left\langle \sum_{i=1}^{N} \frac{(s_i-\mu_i(\boldsymbol{c},\boldsymbol{x};\boldsymbol{c}_0,\boldsymbol{a}))^2}{2\boldsymbol{v}(\boldsymbol{x};\boldsymbol{\sigma})_i^2}\right\rangle_{S+E_n} \\ &- \left\langle \sum_{j=1}^{M} \frac{(c_j-p_j)^2}{2q_j^2} + \right\rangle_{S+E_n} + \left\langle \sum_{j=1}^{M} n_j \ln c_j \right\rangle_{S+E_n}\end{aligned} \tag{C23}$$

$\langle \ln \mathcal{N}_{S+E_n}\rangle_{S+E_n} = \ln \mathcal{N}_{S+E_n}$ since it is a constant. $\langle \ln p(\boldsymbol{x})\rangle_{S+E_n}$ may be computed exactly via numerical summation since $p(\boldsymbol{x})$ is a discrete probability distribution. The quadratic terms may be expanded as,

$$\sum_{j=1}^{M} \frac{(c_j-p_j)^2}{2q_j^2} = \sum_{j=1}^{M} \frac{1}{2q_j^2}(c_j^2 - 2p_jc_j + p_j^2) \tag{C24}$$

and as,

$$\begin{aligned}&\sum_{i=1}^{N} \frac{(s_i-\mu_i(\boldsymbol{c},\boldsymbol{x};\boldsymbol{c}_0,\boldsymbol{a}))^2}{2\boldsymbol{v}(\boldsymbol{x};\boldsymbol{\sigma})_i^2} = \\ &\sum_{i=1}^{N} \frac{1}{2\boldsymbol{v}(\boldsymbol{x};\boldsymbol{\sigma})_i^2}(s_i^2 - 2s_i\mu_i(\boldsymbol{c},\boldsymbol{x};\boldsymbol{c}_0,\boldsymbol{a}) + \mu_i(\boldsymbol{c},\boldsymbol{x};\boldsymbol{c}_0,\boldsymbol{a})^2)\end{aligned} \tag{C25}$$

where $\mu_i(\boldsymbol{c},\boldsymbol{x};\boldsymbol{c}_0,\boldsymbol{a}) = \sum_j a_{ij}x_jc_j + \sum_{j'} a_{ij'}c_{0j'}$. Exact solutions may be found for these terms using the indefinite integral result given by eqn. (C16).

C.6 Derivation of $\left\langle \sum_{j=1}^{M} n_j \ln c_j \right\rangle_{S+E_n}$

The most technically difficult term in eqn. (C22) is $\left\langle \sum_{j=1}^{M} n_j \ln c_j \right\rangle_{S+E_n}$, due to the challenges associated with integrating the $\ln c_j$ terms. For the sake of simplicity, a single logarithmic term, which may be used to derive all other such terms, is considered here. It is given by,

$$\int_0^\infty dc \ln(c)\, c^n e^{-\frac{(c-p)^2}{2q^2}} \tag{C26}$$

We then use a method known within the physics community as the "replica trick" to change the logarithm function in eqn. (C26) into a continuous moment problem.

$$\begin{aligned}
&\int_0^\infty dc \ln(c) c^n e^{-\frac{(c-p)^2}{2q^2}} = \lim_{s\to 0^+} \int_0^\infty dc \frac{e^{s\ln(c)}-1}{s} c^n e^{-\frac{(c-p)^2}{2q^2}} \\
&= \lim_{s\to 0^+} \int_0^\infty dc \frac{c^s-1}{s} c^n e^{-\frac{(c-p)^2}{2q^2}} = \lim_{s\to 0^+} \int_0^\infty dc \frac{c^{s+n}-c^n}{s} e^{-\frac{(c-p)^2}{2q^2}} \\
&= \left[\frac{\partial}{\partial s} \int_0^\infty dc\, c^{s+n} e^{-\frac{(c-p)^2}{2q^2}}\right]_{s\to 0^+}
\end{aligned} \tag{C27}$$

The final term of eqn. (C27) may be interpreted as the derivative evaluation of a Mellin transform-like quantity of the off-center Gaussian function. The Mellin transform itself is defined by

$$\{\mathscr{M}(f)\}(z) = \Phi(z) = \int_0^\infty dx\, f(x) x^{z-1}, \tag{C28}$$

and the Mellin transform of the off-center Gaussian is given by,

$$\{\mathscr{M}(e^{-ax^2+bx})\}(z) = (2a)^{-z/2}\Gamma(z) e^{b^2/(8a)} D_{-z}\left(\frac{-b}{\sqrt{2a}}\right) \qquad \mathrm{Re}(z) > 0 \tag{C29}$$

where $D_\alpha(y)$ is the parabolic cylinder function [79].

Setting $z \to s+n+1$ in eqn. (C29), we find that eqn. (C27) becomes

$$\left[\frac{\partial}{\partial s}\int_0^\infty dc\, c^{s+n} e^{-\frac{(c-p)^2}{2q^2}}\right]_{s\to 0^+} =$$

$$e^{-p^2/(4q^2)}\frac{\partial}{\partial s}\left[q^{(s+n+1)}\Gamma(s+n+1)D_{-s-n-1}\left(\frac{-p}{q}\right)\right]_{s\to 0^+} =$$
$$e^{-p^2/(4q^2)}\left[\ln(q)q^{(n+1)}\Gamma(n+1)D_{-n-1}\left(\frac{-p}{q}\right)+\right.$$
$$\left.q^{(n+1)}\Gamma'(n+1)D_{-n-1}\left(\frac{-p}{q}\right)\right]+$$
$$e^{-p^2/(4q^2)}q^{(n+1)}\Gamma(n+1)\left[\frac{\partial}{\partial s}D_{-s-n-1}\left(\frac{-p}{q}\right)\right]_{s\to 0^+} \tag{C30}$$

where $n \in \mathbb{Z}^{0^+}$, $\Gamma'(n+1) = n!\left(-\gamma+\sum_{k=1}^{n}\frac{1}{k}\right)$ when n is an integer, $\Gamma'(n+1) =$ and γ is the Euler-Mascheroni constant which is approximately equal to 0.577215664901532….

The final term in eqn. (C38), $\left[\frac{\partial}{\partial s}D_{-s-n-1}\left(\frac{-p}{q}\right)\right]_{s\to 0^+}$, requires more careful consideration. The parabolic cylinder function $D_\alpha(y)$ is given by a confluent hypergeometric function [80]

$$D_\alpha(y) = 2^{\alpha/2}e^{-y^2/4}\left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1-\alpha}{2})}{}_1F_1(-\frac{\alpha}{2};\frac{1}{2};\frac{y^2}{2})\right.$$
$$\left.+\frac{y}{\sqrt{2}}\frac{\Gamma(-\frac{1}{2})}{\Gamma(-\frac{\alpha}{2})}{}_1F_1(\frac{1-\alpha}{2};\frac{3}{2};\frac{y^2}{2})\right] \tag{C31}$$

The evaluation of the derivative is performed term-by-term; beginning with the following:

$$\left[\frac{\partial}{\partial s}2^{-(s+n+1)/2}\right]_{s\to 0^+} = \left[-\frac{1}{2}\ln(2)\,2^{-(s+n+1)/2}\right]_{s\to 0^+}$$
$$= -\frac{1}{2}\ln(2)\,2^{-(n+1)/2} \tag{C32}$$

and

$$\left[\frac{\partial}{\partial s}\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{2+s+n}{2})}\right]_{s\to 0^+} = \left[-\frac{\Gamma(\frac{1}{2})\Gamma'(\frac{2+s+n}{2})}{2\Gamma(\frac{2+s+n}{2})^2}\right]_{s\to 0^+}$$
$$= -\frac{\Gamma(\frac{1}{2})\Gamma'(1+\frac{n}{2})}{2\Gamma(1+\frac{n}{2})^2} \tag{C33}$$

and

$$\left[\frac{\partial}{\partial s}\frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{1+s+n}{2})}\right]_{s\to 0^+} = \left[-\frac{\Gamma(-\frac{1}{2})\Gamma'(\frac{1+s+n}{2})}{2\Gamma(\frac{1+s+n}{2})^2}\right]_{s\to 0^+} = -\frac{\Gamma(-\frac{1}{2})\Gamma'(\frac{1+n}{2})}{2\Gamma(\frac{1+n}{2})^2} \tag{C34}$$

A general expression for the derivative of the hypergeometric function portion of eqn. (C31) is expressed as

$$\frac{\partial}{\partial a}{}_1F_1(a;b;z) = \frac{z}{b}F^{1:2;1}_{2:1;1}\left[\begin{matrix}a+1:a,1;1\\ b+1,2:a+1;1\end{matrix}\middle| z,z\right] \tag{C35}$$

which is derived and detailed in [81, 82]. The derivative evaluations for our specific hypergeometric functions are

$$\begin{aligned}&\frac{\partial}{\partial s}{}_1F_1\left(\frac{1+s+n}{2};\frac{1}{2};\frac{p^2}{2q^2}\right)\Bigg|_{s\to 0^+} =\\ &\frac{p^2}{2q^2}F^{1;2;1}_{2;1;1}\left[\begin{matrix}\frac{3+s+n}{2}:\frac{1+s+n}{2},1;1\\ \frac{3}{2},2:\frac{3+s+n}{2};1\end{matrix}\middle|\frac{p^2}{2q^2},\frac{p^2}{2q^2}\right]\Bigg|_{s\to 0^+} =\\ &\frac{p^2}{2q^2}F^{1;2;1}_{2;1;1}\left[\begin{matrix}\frac{3+n}{2};\frac{1+n}{2},1;1\\ \frac{3}{2},2;\frac{3+n}{2};1\end{matrix}\middle|\frac{p^2}{2q^2},\frac{p^2}{2q^2}\right]\end{aligned} \tag{C36}$$

and

$$\begin{aligned}&\frac{\partial}{\partial s}{}_1F_1\left(1+\frac{s+n}{2};\frac{3}{2};\frac{p^2}{2q^2}\right)\Bigg|_{s\to 0^+} =\\ &\frac{p^2}{6q^2}F^{1;2;1}_{2;1;1}\left[\begin{matrix}2+\frac{s+n}{2};1+\frac{s+n}{2},1;1\\ \frac{5}{2},2;2+\frac{s+n}{2};1\end{matrix}\middle|\frac{p^2}{2q^2},\frac{p^2}{2q^2}\right]\Bigg|_{s\to 0^+} =\\ &\frac{p^2}{6q^2}F^{1;2;1}_{2;1;1}\left[\begin{matrix}2+\frac{n}{2};1+\frac{n}{2},1;1\\ \frac{5}{2},2;2+\frac{n}{2};1\end{matrix}\middle|\frac{p^2}{2q^2},\frac{p^2}{2q^2}\right]\end{aligned} \tag{C37}$$

Putting the preceding together, we express

$$\frac{\partial}{\partial s}D_{-s-n-1}\left(\frac{-p}{q}\right)\Bigg|_{s\to 0^+} =$$

$$-\frac{1}{2}\ln(2)2^{-(n+1)/2}e^{-p^2/(4q^2)}\left[\frac{\Gamma(\frac{1}{2})}{\Gamma(1+\frac{n}{2})}{}_1F_1(\frac{n+1}{2};\frac{1}{2};\frac{p^2}{2q^2})\right.$$
$$\left.-\frac{p}{\sqrt{2}q}\frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{n+1}{2})}{}_1F_1(1+\frac{n}{2};\frac{3}{2};\frac{p^2}{2q^2})\right]$$

$$+2^{-(n+1)/2}e^{-p^2/(4q^2)}\left[-\frac{\Gamma(\frac{1}{2})\Gamma'(1+\frac{n}{2})}{2\Gamma(1+\frac{n}{2})^2}{}_1F_1(1+\frac{1+n}{2};\frac{1}{2};\frac{p^2}{2q^2})\right.$$
$$\left.+\frac{\Gamma(\frac{1}{2})}{\Gamma(1+\frac{n}{2})}\frac{p^2}{2q^2}F^{1;2;1}_{2:1;1}\left[\begin{matrix}\frac{3+n}{2};\frac{1+n}{2},1;1\\ \frac{3}{2},2;\frac{3+n}{2};1\end{matrix}\middle|\frac{p^2}{2q^2},\frac{p^2}{2q^2}\right]\right]$$

$$+\frac{p}{\sqrt{2}q}\frac{\Gamma(-\frac{1}{2})\Gamma'(\frac{1+n}{2})}{2\Gamma(\frac{1+n}{2})^2}{}_1F_1(\frac{n+1}{2};\frac{1}{2};\frac{p^2}{2q^2})$$
$$\left.-\frac{p^3}{6\sqrt{2}q^3}\frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{1+n}{2})}F^{1;2;1}_{2;1;1}\left[\begin{matrix}2+\frac{n}{2};1+\frac{n}{2},1;1\\ \frac{5}{2},2;2+\frac{n}{2};1\end{matrix}\middle|\frac{p^2}{2q^2},\frac{p^2}{2q^2}\right]\right] \tag{C38}$$

This completes the exact solution of eqn. (C26).

C.7 Derivation of Normalization Factors

In this subsection, we derive the various normalizations of the constituent probability distributions of P_{S+E}. First we consider the normalization associated with the Gaussian sensor response

$$\mathcal{N}_S(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{\sigma}|\boldsymbol{x}) = \prod_{i=1}^{N}\left(\int_{-\infty}^{\infty}ds_i\, e^{-\frac{(s_i-\mu_i)^2}{2\sigma_i^2}}\right) = (2\pi)^{N/2}\sqrt{\prod_{i=1}^{N}\sigma_i^2} \tag{C39}$$

where $\mu_i = \sum_j a_{ij}x_jc_j + \sum_{j'}a_{ij'}c_{j'}$ and $\sigma_i^2 = \sum_j x_j\sigma_{ij}^2 + \sum_{j'}\sigma_{ij'}^2$ where c_j denotes the randomly generated concentrations of the chemical background whose presence is determined by x_j. $c_{j'}$ denotes the target analytes/chemical simulants/chemical knowns that are always present. For Gaussian noise, we find that $\mathcal{N}_S(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{\sigma}|x)$ isn't dependent upon either $\boldsymbol{a}$ or $\boldsymbol{\mu}$, so we rewrite it as,

$$\mathcal{N}_S(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{\sigma}|\boldsymbol{x}) = \mathcal{N}_S(\boldsymbol{\sigma}|\boldsymbol{x}) \tag{C40}$$

The probability density associated with our model of the external environment is,

$$\rho_{E_n}(\boldsymbol{c},\boldsymbol{x};\boldsymbol{p},\boldsymbol{q},\boldsymbol{n}) = \left(\sum_{\boldsymbol{x}\in[\boldsymbol{0},\boldsymbol{1}]^N} p(\boldsymbol{x})\right)\prod_{j=1}^{M}\left(\int_0^\infty dc_j\, c_j^{n_j} e^{-\frac{(c_j-p_j)^2}{2q_j^2}}\right) \tag{C41}$$

The random vectors generated by the probability distribution associated with the density are $\boldsymbol{c}$ and $\boldsymbol{x}$. They denote the concentrations of the chemicals present and the binary presence or absence of those chemicals respectively. $\boldsymbol{p}$, $\boldsymbol{q}$, and $\boldsymbol{n}$ are vectors of external parameters of the densities $\boldsymbol{n}$ are independent integer valued parameters used to set the shape of the densities. $\boldsymbol{p}$ and $\boldsymbol{q}$ are dependent parameters set in terms of $\boldsymbol{n}$ as well as the expected means and variances of the chemicals in the environment.

Solving for normalizations denoted by $\eta_j(p_j,q_j,n_j)$ associated with the concentration generating portion of the probability distribution

$$\begin{aligned}\eta(p,q,n) &= \int_0^\infty dc\, c^n e^{-\frac{(c-p)^2}{2q^2}} = e^{-\frac{p^2}{2q^2}}\int_0^\infty dc\, c^n e^{-\frac{c^2}{2q^2}+\frac{pc}{q^2}} = \\ &e^{-\frac{p^2}{2q^2}}\left[\frac{1}{2}\sqrt{\frac{\pi}{a}}\sum_{k=0}^{n}\binom{n}{k}\frac{\partial^{n-k}}{\partial b^{n-k}}\left[e^{\frac{b^2}{4a}}\right]\cdot\frac{\partial^k}{\partial b^k}\left[\operatorname{erf}\left(\frac{2ac-b}{2\sqrt{a}}\right)\right]\right]_0^\infty\end{aligned} \tag{C42}$$

where $a = \frac{1}{2q^2}$ and $b = \frac{p}{q^2}$.

Combining the preceding equations allows us to express the total normalization $N_{S+E_n}(\boldsymbol{a},\boldsymbol{\sigma},\boldsymbol{p},\boldsymbol{q})$ so that only the summation over the probabilities associated with the binary variables $\boldsymbol{x}$ is left. This discrete calculation may be solved exactly via direct calculation and is expressed as,

$$\begin{aligned}&\mathcal{N}_{S+E_n}(\boldsymbol{\sigma},\boldsymbol{p},\boldsymbol{q},\boldsymbol{n}) = \\ &\left(\sum_{\boldsymbol{x}\in[\boldsymbol{0},\boldsymbol{1}]^N} p(\boldsymbol{x})\prod_{i=1}^{N}\int_{-\infty}^{\infty} ds_i\, e^{-\frac{(s_i-\mu_i)^2}{2\sigma_i^2}}\right)\cdot\prod_{j=1}^{M}\left(\int_0^\infty dc_j\, c_j^n e^{-\frac{(c_j-p_j)^2}{2q_j^2}}\right) = \\ &\left(\sum_{\boldsymbol{x}\in[\boldsymbol{0},\boldsymbol{1}]^N} p(\boldsymbol{x})\mathcal{N}_S(\boldsymbol{\sigma}|\boldsymbol{x})\right)\left(\prod_{j=1}^{M}\eta(p_j,q_j,n_j)\right)\end{aligned} \tag{C43}$$

C.8 Numerically Solving for the Parameters of the External Environment's Probability Distributions

From eqn. (C41), we consider the single component distribution $\frac{\rho(c;p,q,n)}{\eta(p,q,n)} = \frac{1}{\eta(p,q,n)}\left(c^n e^{-\frac{(c-p)^2}{2q^2}}\right)$ which is used to randomly generate the concentration of an associated background interferent. The parameter n

is an integer-valued parameter which is used to set the shape of the distribution. The parameters p and q may then be tuned such that the distribution has the desired mean concentration and variance expected from known experimental data.

As a refresher we recall the expressions for the various terms we will use to tune p and q: The mean concentration is expressed by

$$\langle c \rangle_n(p,q) = \frac{1}{\eta(p,q,n)} \int_0^\infty dc\, c^{n+1} e^{-\frac{(c-p)^2}{2q^2}} \tag{C44}$$

Likewise, the variance of the mean concentration is

$$\sigma_n^2(p,q) = \langle c^2 \rangle_n(p,q) - \langle c \rangle_n^2(p,q) \tag{C45}$$

with the second order moment given by

$$\langle c^2 \rangle_n(p,q) = \frac{1}{\eta(p,q,n)} \int_0^\infty dc\, c^{n+2} e^{-\frac{(c-p)^2}{2q^2}} \tag{C46}$$

All of the integrals given above may be calculated from the indefinite integral given by eqn. (C16). To set p and q, we take experimental values for the concentration and variance of a given environmental chemical interferent and use numerical root solvers such as the bisection method or Brent's algorithm used in SciPy's fsolve function [83]. We then solve the simultaneous nonlinear equations for p and q given by eqns. (C44) and (C45) for p and q.