SIMULATION-BASED DIGITAL TWIN DEVELOPMENT FOR BLOCKCHAIN ENABLED END-TO-END INDUSTRIAL HEMP SUPPLY CHAIN RISK MANAGEMENT

Keqi Wang
Wei Xie
Bo Wang
Jinxiang Pei

Wencen Wu

Mechanical and Industrial Engineering
Northeastern University
360 Huntington Avenue
Boston, MA 02115, USA

Computer Engineering Department
San Jose State University
1 Washington Sq
San Jose, CA 95192, USA

Mike Baker

Qi Zhou

Willamette Valley Assured LLC.
33490 Honah Lea Drive
Lebanon, OR 97355, USA

QuarkChain Inc.
55E 3rd Ave
San Mateo, CA 94401, USA

ABSTRACT

With the passage of the 2018 U.S. Farm Bill, Industrial Hemp production is moved from limited pilot programs to a regulated agriculture production system. However, Industrial Hemp Supply Chain (IHSC) faces critical challenges, including high complexity and variability, very limited production knowledge, and lack of data/information tracking. In this paper, we propose blockchain-enabled IHSC and develop a preliminary simulation-based digital twin for this distributed cyber-physical system (CPS) to support the process learning and risk management. Basically, we develop a two-layer blockchain with proof of authority smart contract, which can track the data and key information, improve the supply chain transparency, and leverage local authorities and state regulators to ensure the quality control verification. Then, we introduce a simulation-based digital twin for IHSC management, which can characterize the process spatial-temporal causal interdependencies and dynamic evolution to guide risk control and decision making. Our empirical study demonstrates the promising performance of proposed platform.

1 INTRODUCTION

The objective of our study is to develop a simulation-based cyber-physical system (CPS) digital twin for blockchain-enabled industrial hemp supply chain (IHSC), which will be utilized to improve the understanding of end-to-end process, guide quality control verification, and accelerate the development of a safe, efficient, reliable and automated supply chain system. With the passage of the 2018 U.S. Farm Bill, IH production is moved from being limited pilot programs administered by state regulatory officials and becomes a public *regulated* agriculture production. The total area licensed for hemp production has increased dramatically from 37,122 (in year 2017) to 310,721 (in year 2019) acres nationally (Sterns 2019). According to the U.S. Department of Agriculture (USDA), the sales are expected to increase from \$25 million in 2020 to more

than $100 million by 2022. However, there was very limited knowledge/experiences and well-developed guidelines on how to navigate the end-to-end IHSC process, including seed selection, fertility, harvesting and handling on an industrial scale.

The two most common products produced from hemp are cannabidiol (CBD for medicinal uses) and tetrahydrocannabinol (THC for recreational drug use). As opposed to the safety of CBD, THC has been reported with unpleasant side-effects including but not limited to anxiety and panic, impaired attention and memory. Considering the major active ingredient in marijuana causing psychoactive effects, the law (2018 U.S. Farm Bill) requires that CBD plants and derivatives should contain no more than 0.3 percent THC on a dry weight basis. Thus, the development of a safe, reliable, sustainable, efficient and automated IHSC plays a critical role in improving this economy and ensuring public health.

The management of IHSC as a cyber-physical system faces critical challenges, including: high complexity/variability, very limited IHSC knowledge, data tampering, scalability, data/information tracking. First, the IH supply chain process is complex and dynamic. The content flow of CBD/THC highly depends on the complex interactions of many physical/human factors, such as seed varieties, soil properties, weather condition, drying and extraction. Second, the data collection cycle is lengthy (about one year), and the existing IH production knowledge is very limited. Third, the data integrity is hardly guaranteed without human intervention. Fourth, how to allocate the limited inspection resources and monitor this fast growing industry is extremely challenging. *Therefore, IH practitioners and government regulators have urgent needs in improving supply chain transparency, developing deep understanding of the end-to-end IHSC, eliminating risks, ensuring CBD product quality, and controlling the THC content flow in the supply chain.*

Driven by these challenges and needs, we propose blockchain-enabled IHSC and create a digital twin of this CPS to support the process surveillance, learning, and risk management. We explore the two-layer blockchain with proof of authority (PoA) and smart contract consensus (Xie et al. 2019), which can efficiently leverage local authorities (e.g., Willamette Valley Assured LLC.) and limited officials inspection resources (e.g., State Regulatory) to improve the supply chain transparency and ensure the data integrity. We further develop a stochastic simulation for the end-to-end blockchain-enabled IHSC from seed selection to the final commercial product (CBD oil). We consider the complex interactions of random input factors and decisions introduced in different steps and model the dynamic evolution of key attributes of IH product (e.g., CBD and THC). It can characterize the spatial-temporal process interdependencies and provide the reliable guidance on risk analysis/control and decision making for IHSC.

This paper is organized as follows. In Section 2, we provide the problem description and introduce the end-to-end blockchain-enabled IHSC. In Section 3, we discuss the two-layer blockchain and PoA smart contract design for process data tracking and information sharing. In Section 4, we create the simulation model for blockchain-enabled IHSC and provide the process risk analysis. After that, we conduct the empirical study in Section 5 and conclude this paper in Section 6.

2 END-TO-END BLOCKCHAIN-ENABLED INDUSTRIAL HEMP SUPPLY CHAIN

Even though the proposed simulation modeling and blockchain-based CPS platform are applicable to general regulated manufacturing and supply chain, in this paper, we focus on the IHSC from seed selection to final commercial product, called "THC free broad-spectrum CBD oil", which will be shortened as *CBD oil* for simplification. The main operation units of IHSC include: (1) cultivation, (2) stabilization (drying), and (3) manufacturing (extraction, winterization, and purification-PLC); see Figure 1. Since farmers have strong incentive to maintain the lot uniformity, the same seed variety and growing strategy are typically applied to the same lot. Thus, we ignore the IH plant variation in the same lot, and consider the lot-to-lot variation of IHSC outputs, including productivity, quality, and cycle time.

__Cultivation:__ It starts with selection of seed variety, and licensed growers/farmers select one or a few IH seed varieties to grow. Here, we consider one particular variety: Special Sauce provided by Oregon CBD (2020). Once receiving the selected seed, the farmer simultaneously starts Germination and Pre-Plant Soil Preparation process. Germination happens at the greenhouse with the controlled-growth environment, where the seed germinates in about 5–10 days. After germination, there is at most 2-day holding period before
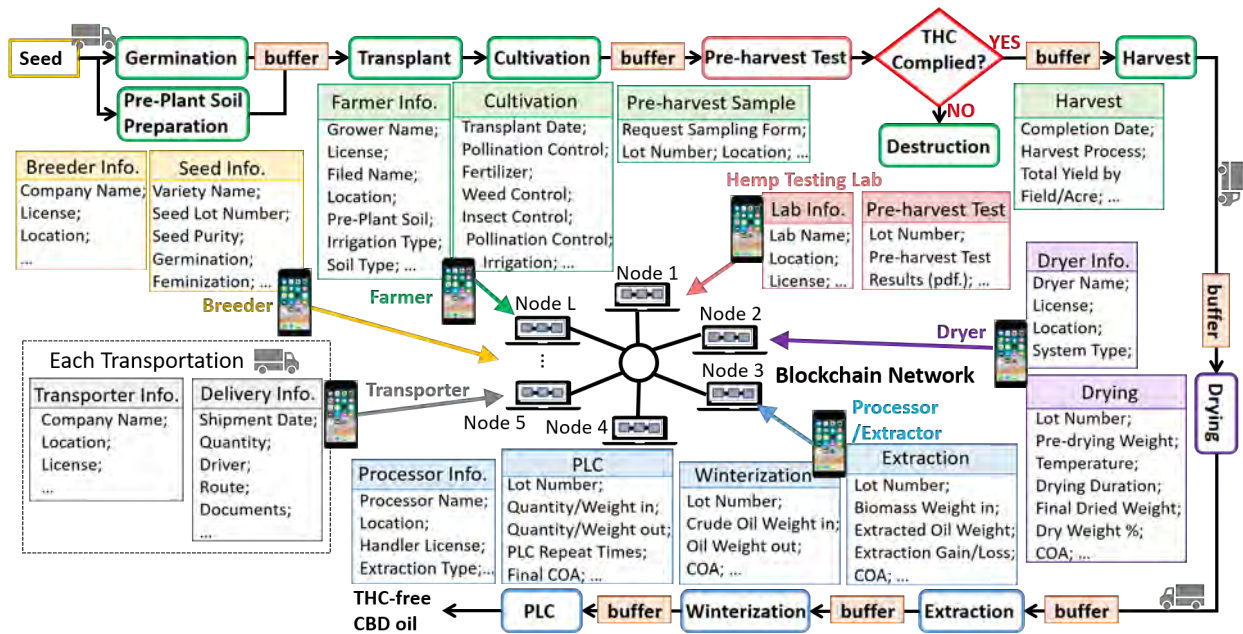
Figure 1: Blockchain-enabled end-to-end industrial hemp supply chain.

transplanting. Otherwise, it is abandoned. It is important that the seedlings are not stressed after leaving the greenhouse and being planted. At the meantime, after the Soil Test, the farmer performs Pre-Plant Soil Preparation to adjust the properties of field soil, which usually takes 1–2 days. Once both steps are finished, the germinated seedlings are transplanted into the field, which takes about 1–2 days.

Then, the farmer continues the cultivation process including two typical sub-phases: Vegetative and Flowering. The Vegetative is the main growth stage, where the iconic fan leaves begin to develop and grow larger with more blades and fuller sets of leaves. The Flowering is the final stage of growth, where the resinous buds develop (the majority of the CBD is generated). Once the flowers begin to develop, they require a different kind of care. Besides the regular control of weed, insect and mold, the most important one is pollination control. If the plants are pollinated, they shift their energy from producing unpollinated flowers that are high in oil to pollinated flowers that consume oil to produce seed, which will decrease the CBD yield significantly. Overall, the cultivation process takes about 50–60 days.

After that, the sample of IH is sent to Hemp Testing Lab for *pre-harvest test* that takes 2–7 days on average and gives the comprehensive test results including CBD/THC levels. *Based on the USDA regulation, the lot with THC level greater than 0.3% is required to be destroyed completely.* For IH lots that have passed the pre-harvest test, farmers can start the harvest process. Usually farmers tend to wait for several days since the CBD level further increases during this period. However, since the THC content also increases, the government requires farmers to complete harvest no longer than 15 days after the pre-harvest test. If farmer doesn't finish harvesting all the lots in time, the rest needs to be scheduled for another testing and then followed by new harvest within another 15-day limit. Thus, farmers need to make an appropriate harvest schedule to consider risk-profit trade-off.

Stabilization: The harvested IH biomass often has a high moisture content, even up to 85%, which has to be reduced to an ideal level to avoid deterioration in storage and transportation. Typically, there are two types of stabilization: drying-stabilizing and freezing-stabilizing. Since drying is most common for the post-harvest stabilization process, we consider it here. Based on the drying machine availability, the harvested IH biomass may have to wait. The probability of being contaminated increases during waiting with high moisture, which can cause lost in profit. The drying process typically takes 1-2 days.

Manufacturing: The dried IH biomass is then sold to licensed processors/extractors to produce the desired commercial product (CBD oil). The main steps include Extraction, Winterization, and Preparatory Liquid Chromatography (PLC). Since the dried biomass has life of 9–12 month, the product can wait in buffer before each step. Extraction is a process to produce crude oil, which typically includes two steps: extraction and decarboxylation. The Ethanol Extraction is one of the most efficient methods for processing large batches of IH biomass. This method cycles ethanol solvent through the solid biomass, strips the cannabinoids, terpenes and plant waxes from the flower to produce the CBD crude oil. Although CBD is the main cannabinoids with the healthy and medicinal potential, the highest concentrations of cannabinoid in biomass (flowers) is cannabidiolic aid (CBDA). After extraction process, the decarboxylation reaction is needed to transform CBDA into its neutral cannabinoids CBD (Wang et al. 2016). Also, the decarboxylation can prevent the degradation of desirable cannabinoids (Romano and Hazekamp 2013). Thus, the oil is to be processed under controlled temperature to decarboxylate the CBDA to CBD. In the mean time, the tetrahydrocannabinolic acid (THCA) is transformed as THC. Not only the cannabinoids but also other undesirable element in biomass are extracted. The whole Extraction process usually takes 1–2 days.

Then, another process, called winterization, needs to be applied to remove the undesirable fat, terpene and wax from crude oil. During this process, the content of the CBD/THC stays almost the same with very minor drop in absolute weight, and the percentage of CBD/THC increases. This process takes about 1–3 days. Finally, several PLC steps are used to remove some specific target molecule, such as THC, and remediate the oil with THC level below the standardized requirement. Considering the economically feasibility, we assume at most twice of this purity process in the empirical study. After the PLC steps, there will be another Certificate of Analysis (COA) test, and the final product can only be acceptable at less than 0.05% THC (for THC free broad-spectrum CBD oil).

3   BLOCKCHAIN BASED DATA/INFORMATION TRACKING SYSTEM

Here we present the two-layer blockchain based data tracking, information sharing, and interoperability framework for the end-to-end IHSC (see Figure 1), which can greatly improve both security and efficiency. In Section 3.1, we first describe the important data/information collected by each group of participants. Here, we consider a static setting with fixed participants. In Section 3.2, we review the two-layer blockchain design and the PoA based smart contract proposed in our previous study (Xie et al. 2019), and then present how to apply it to the end-to-end IHSC to support interoperability and improve safety and efficeincy.

3.1 Participants of IHSC and Data Collection

Here, we list main participants of end-to-end IHSC. By tracking all the critical information/data provided by each participant, we are able to keep the comprehensive monitoring over the process.

Breeder: A breeder produces one or several IH seed varieties or clones, which will be sold to the farmers. Prior to planting and breeding, breeders are required to register and provide the information/data as follows.
(1) *Breeder registration information*: breeder name, registration, certificate, location and so on. This information provides a unique reference to each breeder.
(2) *Seed type/source*: variety name and description, seed lot number, seed purity analysis, pure seed percentage, seed germination percentage, complete seed germination data, flowering type, ferminization process, seed ferminization percentage, and clone information. The seed type/source data will be included in each transaction data when a farmer purchases any seed/clone variety from a breeder.

Licensed Grower: A grower obtains the seeds varieties from breeder, and then cultivate the IH until harvest. Before growing IH, farmers are required to have certificate and provide the information as follows.
(1) *Grower information*: The name and license information provide a unique reference to each grower.
(2) *Field information*: field name, GPS field location, approved background check, pre-plant soil tests, irrigation type, prior year field history, and soil type.
(3) *Cultivation data*: seeding/transplanting data, plant density, row width, irrigation type/frequency/volume, fertilizer frequency/volume, weed control, insect control, mold control, pollination control. During the

germination, transplant and cultivation process, the critical cultivation data, i.e., field/IH plant pictures/image data will be taken timely and uploaded to keep a record of the growing process for each lot.

(4) *Pre-harvest test request*: Pre-harvest Hemp Sampling and Testing Request Form. The growers fulfill the document and send it to the approved lab requesting a lot to be sampled. Typically, this form includes lot number, GPS location and acres.

(5) *Harvest data*: harvest process and completion dates, moisture content, total yield by acre. After receiving the pre-harvest test results, following the 0.3% THC requirement, the grower will either destroy or start to harvest the corresponding IH lots and upload the harvest information to blockchain. *At this stage, the data provided by growers has to be verified and confirmed by local authorities and regulatory officials.*

Dryer: A dryer provides the harvested IH drying-stabilization service for the growers. The dryers need to provide the drying-stabilization data and also post stabilization test results.

(1) *Drying-stabilization data*: dryer name/location/license, type of drying system, lot number, pre-drying weight, temperature/duration of drying, final dried weight, dry weight, container type/weight, brushing to remove excess stalks. It provides the drying process record for each lot of IH.

(2) *Post stabilization test*: cannabinoid content (CBD and THC), pesticide residue, heavy metals. The COA test is needed for trading the biomass to an extractor or the crop analysis records for vertically integrated organizations. At this stage, THC can be higher than 0.3% and the biomass must be tracked.

Licensed Processor/Extractor: A processor purchases the dried IH biomass from the growers/dryers, and then conduct the processing steps to produce commercial CBD oil. He needs to provide the extraction data, winterization data, PLC data, and the post PLC COA test results. *After the post PLC COA test, the data/information provided by the processors needs to be verified and confirmed by local authorities and regulatory officials.* The verified data will be included in new blocks and attached to the blockchain.

(1) *Extraction data*: processor-extractor name/location, handler license information, extraction type, lot number, biomass weight in, extraction input quantity/recaptured, quantity of oil extracted, gain/loss of extraction, post extraction test. The extraction data provides a record of processor/handler information (i.e., responsible operator) and extraction process information.

(2) *Winterization data*: lot number, crude oil weight in, winterized oil out, and post winterization test. The processors need to provide the record of the winterization process on the crude oil.

(3) *PLC data*: lot number, quantity/weight in, quantity/weight out, PLC repeat times, and post PLC test. After the PLC COA test, all the data collected from the processor need to be verified by authorities and further uploaded to the blockchain. The finished commercial CBD oil, with the comprehensive and reliable record of the end-to-end IHSC, can enter the consumer market.

Transporter: A transporter delivers the items (intermediate IH product) between different participants of the IHSC. He needs to provide the transportation data, including sender, receiver, and delivery information.

(1) *Transportation data*: transporter information, driver license, lot number, sender/receiver information, quantity, shipment/delivery date, vehicle/model of transport, route of transportation. By tracking the transportation data, the system keeps the record of comprehensive delivery information for each product/IH biomass. In addition, drivers will use the corresponding delivery information as the legal permission for carrying IH product under road investigation.

Hemp Testing Lab: A Hemp Testing Lab receives the pre-harvest IH samples from the official sampler and then provide the pre-harvest test results and send back to the growers.

(1) *Lab information*: Lab name, location, and other related information, which will provide a unique reference to each lab.

(2) *Pre-harvest test*: sampling date, lot number, COA test, cannabinoid content (CBD and THC), pesticide residue, heavy metals, etc. The official sampler will take the sample from the field after the Lab receiving the request form. He completes the Pre-Harvest Hemp On-site Sampling form and sends the sample to the Lab for analysis. Then the Hemp Testing Lab will upload the pre-harvest results for each lot sample to blockchain, and send the copy back to the grower.

Authorities/Regulators: The local authorities are responsible for the verification in their local areas, and then the regulators (i.e., state USDA) will confirm all the uploaded data (may randomly select some for detailed verification). Regulatory officials will upload the confirmed data and decision to the blockchain. *Suppose the onsite verification takes much more time than online confirmation.*

(1) *Authorities information*: the company information, location, certificates, and other information, which will provide a unique reference to each group of local authorities.

(2) *Verification/confirmation results*: the local authorities and regulatory officials will provide the verification and confirmation decision, as well as the date/time, and responsible persons or group.

3.2 Two-Layer Blockchain to Support IHSC Interoperability and Safety

Here, we briefly review the two-layer blockchain and the PoA based smart contract proposed in our previous study (Xie et al. 2019), and then discuss how to extend it to integrated IHSC. The blockchain is a distributed database, also a global ledger that records all the process data in the network as a timestamp chain of blocks. *Each block contains three parts: header, data body, and validation;* see Figure 2. More specific, the header records the block created time, the height of the current block, the "Nonce" number used to encipher the hashed block, and also the "Merkle Root" that saves the hash of several data/documents (e.g., image or pdf files) and validation information. Each block references to the hash of the block that comes before it, which establishes a chain of the blocks (called blockchain). When new valid data/information records need to be written in the global ledger, new blocks will be generated and added to the blockchain, and broadcast to the network so that each node will update the copy of the ledger. Therefore, the blockchain technology provides the data tracking and information sharing, which can fulfill the needs for IHSC, including transparency, consistency, interoperability, safety, and data integrity.

To facilitate the automation and support the participants reaching a common global view of the world state, each blockchain network needs to establish consensus rules that each data record (or transaction) should conform to. It can be achieved by smart contract (Bahga and Madisetti 2016). Any participant can send new data/information record to the contract for the verification, and then the verified data/information is broadcast to each node in the blockchain network to reach global consensus. For the highly regulated IHSC, Proof-of-Authority (PoA) smart contract is used to ensure that the blockchain contains valid and reliable data/information and reduce the impact of data tampering and cyberattacks. Specifically, each record will be assigned by smart contract to an authority node (e.g., state regulators) for verification, and then the verified data are uploaded to blockchain. In our project, on-site investigations are required to ensure that the data records of IHSC (i.e., images and reports) are accurate and reliable.

However, for typical one-layer blockchain structure and current IHSC design with only state regulators in charge of inspection, the verification efficiency totally depends on the number of available officials for on-site visit. Given the limited inspection resources at state USDA, it is challenging to handle the situation when the scale of the IH market grows dramatically. *Therefore, we consider the two-layer blockchain platform which can leverage the additional resources from local authorities with state regulators, and improve the efficiency and security of IHSC; see the structure illustration in Figure 2.* The two-layer blockchain includes: (1) the sharding layer supported by local authorities for verification, which divides the IHSC or IH market into multiple area/location-based shards; and (2) root chain layer supported by state regulators, which serves as the coordinator among all shard chains. We use the root-chain-first PoA which has a *hierarchical verification procedure*. When any process data/information records (e.g., pre-harvest test results, harvest data) need to be uploaded to the blockchain, the smart contract on corresponding shard will send verification request to the local authority, that will then dispatch investigator for on-site validation. The validated data together with investigator's signature will be included in new shard block, which will be attached to the shard chain and broadcast to the network. After that, the state regulator will review and online confirm the new uploaded data/information, which is relatively quick. Once complete confirmation, a new root block including the corresponding shard block header will be generated and append to the root chain, and the data/information is viewed as "authentic" in the IHSC blockchain network. *The state*

sharding technology and root-chain-first consensus enable different local authorities to simultaneously process the data coming into different shards, which can improve the IHSC efficiency and interoperability.
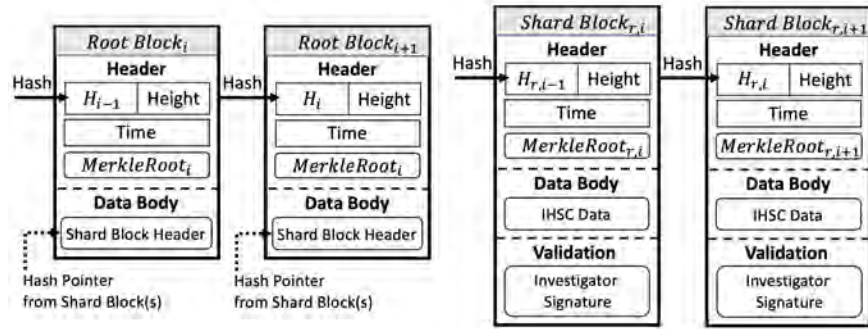

Figure 2: The root (on the left) and shard (on the right) chains structure.

4 SIMULATION-BASED DIGITAL TWIN FOR BLOCKCHAIN-ENABLED IHSC

We develop the simulation model as digital twin of IHSC cyber-physical system, which can characterize the interactions of random input factors and quantify the output variability. It can facilitate the risk management of blockchain-enabled IHSC and guide dynamic decision making. We can track the dynamic evolution of key attributes (i.e., CBD and THC) during the end-to-end IHSC. Since the IH is seasonal plant, we assume all the lots receive the seeds at the same time for each growing season. Thus, we consider the deterministic arrivals of certain lots (say *n*) of seeds, with inter-arrival time equals to the grow cycle (i.e., 365 days).

4.1 Input Modeling

Given limited knowledge and historical data of IHSC process, we develop the input models characterizing the main sources of uncertainty in physical and cyber parts of blockchain-enabled IHSC.

4.1.1 Cultivation Process

After receive the seeds, the farmers will start Germination and Pre-Plant Soil Preparation simultaneously.
Germination: During germination, there's no CBD and THC, we model the germination time following uniform distribution $\text{Unif}(5,10)$ days. Then, the germinated IH seedlings need to wait in buffers until the field is available. If the seedling wait time is more than 2 days, this lot will be dropped.
Pre-Plant Soil Preparation: For pre-plant soil preparation, suppose farmers have already done the soil test and known their soil properties. We model the preparation time following uniform distribution $\text{Unif}(1,2)$.
Transplant: The germinated IH seedlings will be transplanted into the field. We model the transplant time following uniform distribution $\text{Unif}(1,2)$.
Cultivation (Vegetative and Flowering): Due to lack of historical process data, we model Vegetative and Flowering together as the Cultivation process following uniform distribution $\text{Unif}(50,60)$. This is the main process that develops the CBD and THC, and the growth processes are modeled by the linear models,

$$U_c = (gt+\varepsilon)\frac{r}{r+1}, \quad V_c = (gt+\varepsilon)\frac{1}{r+1} \tag{1}$$

where U_c and V_c denote the CBD and THC levels after cultivation step respectively, g is the growth/cumulative rates of total cannabinoid content (CBD + THC) in percentage level (Oregon CBD 2020), t is the flowing time, and r is the ratio of CBD to THC (for the particular seed variety we consider, $r = 28:1$). The growth intrinsic uncertainty $\varepsilon \sim N(0, gt\lambda^2)$ modeled by truncated Normal distribution quantifies the lot-to-lot variation in total cannabinoid (i.e., CBD and THC) level, which is induced by various factors, including weather, soil, moisture, and seed properties, and farmer's growing skill.

Pre-Harvest Test: After the flowering, the official sampler (working for lab) will take samples of each lot and send to Hemp Testing Lab for pre-harvest test. The law requires that uniform samples are taken so that they can represent the true average total THC in the lot. We model the testing time following uniform distribution $\text{Unif}(2,7)$. Based on the equipment availability, the IH samples may wait in the buffer for testing. If certain lot has the THC level greater than the required threshold $\gamma_v = 0.3\%$, the IH in that lot must be destroyed entirely. No harvest can begin until the pre-harvest test results are received.

Harvest: After receiving the pre-harvest test results (and passed the test), the farmer could start the harvest procedure. *During the period from taking the pre-harvest sample to finish harvest, denoted by t', both CBD and THC will continue to increase.* The government requires that the time for completing the harvest can not exceed 15 days, i.e., $t' \leq 15$. If farmer doesn't finish harvesting these lots in time, the rest needs to be scheduled for another testing and harvest later. Therefore, the farmer may need to make an appropriate harvest schedule, balancing the profit from more CBD and the randomness of pre-harvest test and harvest times. We model the harvest process time following uniform distribution $\text{Unif}(1,2)$. The increments of CBD and THC still follow the same linear growth model in Equations (1). So we have, $U_h = U_c + (gt' + \varepsilon')r/(r+1)$ and $V_h = V_c + (gt' + \varepsilon')/(r+1)$, where U_h, V_h denotes the CBD and THC level after harvest step respectively, and the lot-to-lot variation in CBD and THC level is modeled by truncated Normal distribution $\varepsilon' \sim N(0, gt'\lambda^2)$. The harvested IH biomass will be stored in buffer and wait for stabilization (drying). Notice that due to high moisture, too much time spent on waiting for drying will increase the probability of contamination and degradation. In simulation, if the harvested IH biomass waits for more than 3 days before drying, we directly discard this lot of product.

4.1.2 Stabilization Process

For the stabilization process, we mainly consider the drying procedure. The main purpose of drying is to remove the moisture from the harvest IH biomass, and it does not impact on the CBD and THC content. We model the drying time for each lot following uniform distribution $\text{Unif}(1,2)$. We denote the number of drying machines by N_d, which is a process decision parameter. Notice that the dried IH biomass becomes much stable with a lifetime of 9-12 month. Thus, in the following processes, we can ignore the degradation of cannabinoid and contamination of IH biomass during waiting in the buffers. The farmer will sell the dried IH biomass to the processors/extractors, that will start to produce the CBD oil.

4.1.3 Manufacturing Process

After the processors/extractors obtain dried IH biomass from farmers, they will perform the extraction, winterization, and one or two PLC steps in sequence. Before each step, the intermediate IH product can wait in buffers until the corresponding equipment is available for the next step.

Extraction (Extraction and Decarboxylation): Suppose decarboxylation begins right after the extraction. When there is extraction vessel available, we can start the extraction process to obtain the CBD-riched crude oil. We model the extraction time following uniform distribution $\text{Unif}(1,2)$. The mass extracted (i.e., CDB and THC) is a random proportion of the original amount, $U_e = Q \cdot U_h$ and $V_e = Q \cdot V_h$, where $Q \sim \text{Unif}(0.6,0.8)$. The extracted crude oil will wait in the buffer until the winterization equipment available.

Winterization: Since both cannabinoids and waxes present in biomass need to be extracted. The winterization is performed to remove undesirable fat, terpene and waxes from crude oil. We model the winterization process time with uniform distribution $\text{Unif}(1,3)$. During this procedure, we assume the random proportion of CBD/THC is removed, $U_w = W \cdot U_e$ and $V_w = W \cdot V_e$, where $W \sim \text{Unif}(0.95,1)$. The CBD oil after winterization will wait in the buffer until the PLC equipment available.

Preparatory Liquid Chromatography (PLC): Since our target product is THC free broad-spectrum CBD oil, one or two PLC steps are applied to remove THC, while trying to keep CBD. We model each of the PLC process time following uniform distribution $\text{Unif}(1,5)$. Each step of PLC removes the random proportion of CBD/THC, $U_f = Q_u \cdot U_w$ and $V_f = Q_v \cdot V_w$, where $Q_u \sim \text{Unif}(0.9,1.0)$, and $Q_v \sim \text{Unif}(0.3,0.5)$. Notice after the PLC steps, a COA test needs to be performed on the finished CBD oil. The final product can only be acceptable if the percentage of THC is less than the requirement $\gamma = 0.05\%$.

4.1.4 Tracking and Verification through Blockchain Network

To develop the digital twin of blockchain-enabled IHSC, we further create the simulation for the blockchain network. Suppose each participant corresponds to a node N_i of the blockchain network; see the illustration in Figure 1. When any node (e.g., farmer, processor) sends the data/transaction to the smart contract for verification, the smart contract will assign it to local authorities and wait for verification results. After verification, a new block including the verified data/transaction will be generated and appended to the shard chain. The verified data/transaction then requires further (but quick) confirmation/documentation by a state USDA regulator. After that, a new root block including corresponding shard block header will be generated and attached to the root chain. At current stage, the regulated IHSC requires on-site verification by authorities, which will be partially replaced by automatic verification in our future work, through the knowledge model learned from historical data. Compared to the time of verification and confirmation, the delay in network transmission is negligible, which will be ignored for simplification.

Verification: For each shard chain having one local authorities group, we model the verification process as a multiple servers queue, with the number of resources (i.e., officers for verification) denoted by n_s. We model the verification time following exponential distribution $\text{Exp}(\mu_v)$, with mean $\mu_v = 0.1$ day.

Confirmation: Similarly, we also model the confirmation process in root chain as a multiple servers queue, with the number of resources (i.e., state USDA regulators for confirmation) denoted by n_r. We model the confirmation time following exponential distribution $\text{Exp}(\mu_c)$, with mean $\mu_c = 0.05$ day.

4.2 End-to-End IHSC Process Uncertainty Quantification and Risk Analysis

The computer simulation developed as digital twin of real blockchain-enabled end-to-end IHSC is utilized to conduct the process risk analysis and guide the decisions to improve the system safety and efficiency. The simulation output of IHSC depends on the complex interactions of various random inputs. Suppose that there are L sources of uncertainty characterized by the input models, $F = \{F_1, \ldots, F_L\}$, where F_ℓ with $\ell = 1, 2, \ldots, L$ quantifies the uncertainty from physical/cyber parts of blockchain-enabled IHSC. Thus, the vector $\mathbf{Z} = \{Z_1, \ldots, Z_L\}$, representing the collection of all random inputs, can monitor the IHSC process and impact on the outputs variability, where $Z_\ell \sim F_\ell$. Denote the decisions by $\boldsymbol{x} = (x_1, x_2, \ldots, x_K)^\top$, i.e., harvest/drying scheduling and amount of resources allocated at each operation unit. Then, the simulation output can be written as a function of decisions and randoms inputs, denoted by $Y(\boldsymbol{x}; \mathbf{Z})$, such as final CBD oil percentage, THC level, and cycle time of each lot. The lot-to-lot output variation is characterized by variance $\text{Var}[Y(\boldsymbol{x}; \mathbf{Z})]$, which can be estimated by the simulation outputs $\{Y(\boldsymbol{x}; \boldsymbol{z}^{(1)}), \ldots, Y(\boldsymbol{x}; \boldsymbol{z}^{(n)})\}$,

$$\widehat{\text{Var}}[Y(\boldsymbol{x}; \mathbf{Z})] = \frac{1}{n-1} \sum_{i=1}^{n} [Y(\boldsymbol{x}; \boldsymbol{z}^{(i)}) - \bar{Y}]^2, \quad \text{with } \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y(\boldsymbol{x}; \boldsymbol{z}^{(i)}).$$

We are interested in quantifying the contribution of each random input Z_ℓ to output variation. Here we explore the Shapley value (SV) based variance decomposition for IHSC risk analysis, which is motivated by game theory; see Song et al. (2016), Owen (2014). The contribution from Z_ℓ can be measured by

$$s_\ell = \sum_{\mathscr{J} \subset \mathscr{L}/\{\ell\}} \frac{(L - |\mathscr{J}| - 1)! |\mathscr{J}|!}{L!} [c(\mathscr{J} \cup \{\ell\}) - c(\mathscr{J})], \qquad (2)$$

where $\mathscr{L} = \{1, \ldots, L\}$ denotes the index set of random inputs, $|\cdot|$ indicates the set size, and cost function $c(\mathscr{J}) = \text{E}[\text{Var}[Y(\boldsymbol{x}; \mathbf{Z})|\mathbf{Z}_{-\mathscr{J}}]]$ for any subset $\mathscr{J} \subset \mathscr{L}$, and $\mathbf{Z}_{-\mathscr{J}} = \mathbf{Z}_{\mathscr{L}/\mathscr{J}}$ denotes the remaining inputs. The SV in Equation (2) has the property that $\text{Var}[Y(\boldsymbol{x}; \mathbf{Z})] = \sum_{\ell=1}^{L} s_\ell$. Therefore, it can be interpreted as the uncertainty/risk contribution from the ℓ-th random input, which can guide end-to-end IHSC risk control.

Since we could not compute the analytical conditional variance and mean in cost function, we estimate $c(\mathscr{J})$ by simulation samples by using the procedure as follows.

1. For $k = 1, 2, \ldots, K$,

(1) generate the k-th outer sample of random inputs $\mathbf{z}^{(k)}_{-\mathscr{J}}$,

(2) For $i = 1, 2, \ldots, I$, generate the i-th inner sample of random inputs $\mathbf{z}^{(i)}_{\mathscr{J}}$, and obtain simulation output realization $Y(\mathbf{x}; \mathbf{z}^{(i)}_{\mathscr{J}}, \mathbf{z}^{(k)}_{-\mathscr{J}})$

2. Then, the cost function $c(\mathscr{J}) = \mathrm{E}[\mathrm{Var}[Y(\mathbf{x}; \mathbf{Z})|\mathbf{Z}_{-\mathscr{J}}]]$ can be estimated by,

$$\widehat{c}(\mathscr{J}) = \frac{1}{K(I-1)} \sum_{k=1}^{K} \sum_{i=1}^{I} \left[Y(\mathbf{x}; \mathbf{z}^{(i)}_{\mathscr{J}}, \mathbf{z}^{(k)}_{-\mathscr{J}}) - \bar{Y}^{(k)} \right]^2, \quad \text{with} \quad \bar{Y}^{(k)} = \frac{1}{I} \sum_{i=1}^{I} Y(\mathbf{x}; \mathbf{z}^{(i)}_{\mathscr{J}}, \mathbf{z}^{(k)}_{-\mathscr{J}}). \quad (3)$$

By plugging $\widehat{c}(\mathscr{J})$ by eq. (3) into eq. (2), we can obtain the estimation of SV. An efficient approximation algorithm (which uses the randomly selected subset instead of all possible index sets permutations) can be used to reduce the computational burden; see Song et al. (2016).

5 EMPIRICAL STUDY

Here, we conduct the simulation experiments to study the performance of blockchain-based IHSC in terms of improving safety, efficiency, transparency, and reliability. We use the fixed number of resources at each step: (1) for tranplant and harvest, $n_f = 10$; (2) for pre-harvest, $n_l = 10$; (3) for drying, $n_d = 3$; (4) for extraction, winterization and PLC, $n_p = 2$. We set the number of lots to be $n = 50$. The growth rate of cannabinoid is calculated as $C_8/56 = 10.2\%/56 = 0.18\%$ per day, where C_8 repsents the cannabinoid content 8 weeks (56 days) after flowering, and we set $\lambda = 50\%$ (Oregon CBD 2020). For simplification, we assume the limit of waiting time for transplant L_t and dry L_d as 2 days. To avoid the product destruction required by the regulation, suppose each participant could tamper data with probability $p_2 = 30\%$. For the two-layer blockchain, let the verification and confirmation times follow the exponential distributions, i.e., $F_v \sim exp(\mu_v)$ and $F_c \sim \exp(\mu_c)$ with means $\mu_v = 0.1$ and $\mu_c = 0.05$. There are 2 shards with $n_s = 4$ local validators separately and two state regulators $n_r = 2$ in charge of confirmation at the root chain.

5.1 Study the Performance of Blockchain-Enabled IHSC

In this section, we have the simulation experiments to study the performance of IHSC. We set the simulation warmup and run-length to be 200 and $K = 500$, and set the number of replications $J = 100$.

IHSC Security and Safety: *We first show the blockchain-enabled IHSC can improve the safety.* Here we compare the performance of IHSC with and without blockchain. Set the voting panel size $m = 1$. For the critical record (i.e., harvest data, final product quality data), the validators will have onsite visit to ensure the data integrity. To assess the IHSC safety, we record three false pass rates: (1) for pre-harvest test, the expected percentage of lots with more than 0.3% THC that aren't destroyed, $q_{fp} = \lim_{K \to \infty} \mathrm{E}[K_{fp}/K]$; (2) for the 15-day harvest regulation, the expected percentage of lots completed harvest violating the 15 days requirement, $q_{fh} = \lim_{K \to \infty} \mathrm{E}[K_{fh}/K]$; and (3) the expected percentage of lots with more than 0.05% THC and reaching to customers, $q_{fq} = \lim_{K \to \infty} \mathrm{E}[K_{fq}/K]$, where K_{fp}, K_{fh}, K_{fq} are the counts of corresponding false approval. The simulation results in Table 1 illustrate that blockchain can greatly improve the IHSC safety.

Table 1: Simulation results of security improvement.

Security	With Blockchain	Without Blockchain
False Pass Pre-Harvest	0±0%	2.03±1.23%
False Pass Harvest	0±0%	2.98±2.28%
Fake Qualified	0±0%	0.68±0.69%

Efficiency of Interoperability and Data Verification: *Two-layer blockchain can leverage local authorities and state regulatory inspection resources to improve the IHSC verification efficiency and facilitate the scale-up.* Suppose the amount of state regulatory official resources assigned for verification is fixed. We compare the traditional single chain with our two-layer chain with 2 shards. For the single chain, the verification time for each record follows the distribution, $F_s \sim exp(\mu_s)$ with mean $\mu_s = 0.15$. In Table 2, we record

the expected verification/confirmation time for each lot. The results show that two-layer blockchain can improve the efficiency. In addition, the expected number of dropped lots due to too long waiting time for dry, also decreases. There are considerable lots of biomass that need to wait for verification and cannot be dried in time, which causes the loss. Thus, two-layer blockchain can improve the efficiency, reduce the cycle time and eliminate the economy loss.

Table 2: Simulation results of cyber efficiency improvement.

Scalability	Two-Layer Blockchain	Classical Blockchain
Mean Verification Time	0.56±0.03	0.85±0.1
SD Verification Time	0.26±0.17	0.7±0.26
Finished #	275.3±27	242.28±29.91
Dry Drop #	39.42±33.29	66.18±34.34

IHSC Transparency and Process Control: Built on the transparency of IHSC provided by blockchain, every participant can dynamically schedule their resources with more efficiency. We implement the simulations under two situations: (1) the stabilizers don't know how many lots of biomass will come and fix the resource of dry machine $n_d = 3$; and (2) they can dynamically change the resource to avoid the dry-drop. The simulation results recorded in Table 3 illustrate that blockchain can facilitate the process control, improve the efficiency, and reduce the loss.

Table 3: Simulation results of physical efficiency improvement.

Information Share	Fixed Resource	Dynamic Resource
Finished #	275.3±27	306.11±23.9
Dry Drop #	39.42±33.29	3.04±3.79

5.2 IHSC Process Risk Analysis

Based on the study presented in Section 4.2, we conduct the IHSC process risk analysis to quantify the impact of each random input on the CBD and THC levels of final product. For the CBD oil percentage, the variation comes from $L = 7$ *random inputs* $\mathbf{Z} = \{\varepsilon, t', \varepsilon', Q, W, Q_u, Q_v\}$, including (1) growth intrinsic uncertainty during cultivation ε; (2) the period t' between taking the pre-harvest sample and finishing the harvest; (3) growth intrinsic uncertainty from pre-harvest to harvest ε'; (4) extraction rate Q; (5) winterization rate W; (6) CBD removed rate during PLC Q_u; (7) THC removed rate during PLC Q_v. Particularly, the distribution and variability of t' is estimated by using the simulation. The number of random subset permutations, $m = 3000$, is used to estimate the SV. In addition, the variation of THC level caused by $L = 6$ random inputs $\mathbf{Z} = \{\varepsilon, t', \varepsilon', Q, W, Q_v\}$. Here, we consider all possible permutations $m = L! = 720$. We set the number of outer samples $K = 10$, the number of inner samples $I = 100$ and macro-replications $J = 10$. The simulation results of relative contribution from each input factor are calculated by applying Eq. (2) and (3). The results from J macro-replications, $RC_\ell = \frac{1}{J} \sum_{j=1}^{J} \frac{s_\ell^{(j)}}{\widehat{\text{Var}}_{[Y(\mathbf{x};\mathbf{Z})]}}$, are represented in Table 4. Clearly, the most part of lot-to-lot variation of both CBD and THC comes from the cultivation step.

Table 4: The estimated relative contribution of each random input factor.

Input	ε	t'	ε'	Q	W	Q_u	Q_v
CBD %	69.8±5.3	3.0±1.8	8.0±3.0	4.9±1.5	4.2±1.4	4.4±2.9	5.4±2.8
THC %	60.9±3.2	2.4±1.8	14.8±3.6	6.9±2.6	1.5±2.2	N.A.	15.0±2.0

6 CONCLUSION

In this paper, we propose the blockchain-enabled IHSC and introduce the simulation-based digital twin to: (1) improve the IHSC transparency, safety, security, and efficiency; (2) conduct the process risk analysis and guide the decisions. The empirical results indicate that the proposed platform has promising performance. *For the future research, based on the real data collected in our project, we will focus on: (1) providing deep, interpretable, and data-efficient IHSC process learning (e.g., how the various factors, including the seed selection, soil and weather, impact on the IH production); (2) proposing the knowledge-based participants' reputation learning; and (3) improving and validating the digital twin so that it can provide high fidelity of real-world blockchain-enabled IHSC risk behaviors; and (4) accelerating the end-to-end IHSC automation.*

REFERENCES

Bahga, A., and V. K. Madisetti. 2016. "Blockchain Platform for Industrial Internet of Things". *Journal of Software Engineering and Applications* 9(10):533–546.

Oregon CBD 2020. "Farmers Resources". https://oregoncbdseeds.com/resources/, accessed 25nd August 2020.

Owen, A. B. 2014. "Sobol'indices and Shapley value". *SIAM/ASA Journal on Uncertainty Quantification* 2(1):245–251.

Romano, L. L., and A. Hazekamp. 2013. "Cannabis oil: chemical evaluation of an upcoming cannabis-based medicine". *Cannabinoids* 1(1):1–11.

Song, E., B. L. Nelson, and J. Staum. 2016. "Shapley effects for global sensitivity analysis: Theory and computation". *SIAM/ASA Journal on Uncertainty Quantification* 4(1):1060–1083.

Sterns, J. A. 2019. "Is the Emerging US Hemp Industry Yet Another Boom–Bust Market for US Farmers?". *Choices* 34(316-2019-4152):1–8.

Wang, M., Y.-H. Wang, B. Avula, M. M. Radwan, A. S. Wanas, J. van Antwerp, J. F. Parcher, M. A. ElSohly, and I. A. Khan. 2016. "Decarboxylation study of acidic cannabinoids: a novel approach using ultra-high-performance supercritical fluid chromatography/photodiode array-mass spectrometry". *Cannabis and Cannabinoid Research* 1(1):262–271.

Xie, W., B. Wang, Z. Ye, W. Wu, J. You, and Q. Zhou. 2019. "Simulation-based Blockchain Design to Secure Biopharmaceutical Supply Chain". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 797–808. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

KEQI WANG is a Ph.D. student of the Department of Mechanical and Industrial Engineering (MIE) at Northeastern University (NEU). His research interests are blockchain, stochastic simulation, and data analytics. His email is wang.keq@husky.neu.edu.

WEI XIE is an assistant professor in the Department of MIE at NEU. She received her M.S. and Ph.D. in Industrial Engineering and Management Sciences at Northwestern University. Her research interests are in interpretable artificial intelligence, computer simulation, data analytics, stochastic optimization, and CPS risk management. Her email address is w.xie@northeastern.edu.

WENCEN WU is an assistant professor of the Computer Engineering Department of San Jose State University. She obtained her Ph.D from the Georgia Institute of Technology. Her research interests include autonomous multi-robot systems, cyber-physical systems, and distributed parameter systems. Her email address is wencen.wu@sjsu.edu.

BO WANG is a Ph.D. candidate of the Department of MIE at NEU. His research interests are data analytics, input modeling and uncertainty quantification in stochastic simulation. His email address is wang.bo2@husky.neu.edu.

JINXIANG PEI is an assistant teaching professor in the Department of MIE at NEU. He received his M.S. and Ph.D. in Industrial Engineering and Management Sciences at Northwestern University. His research interests are machine learning and supply chain management. His email address is ji.pei@northeastern.edu.

MIKE BAKER is the Founder of Willamette Valley Assured LLC., focusing on quality control for the entire process of industrial hemp. His email address is mbaker@wvassured.com.

QI ZHOU is the Founder and CEO of QuarkChain Inc., a blockchain company providing flexible, scalable, and user-oriented blockchain solutions. He received his Ph.D. from the Georgia Institute of Technology. His email address is qizhou@quarkchain.org.