HAPI Whitepaper 2.0

Created by: HAPI Team

HAPI is a set of cross chain smart contracts that are embedded into DeFI products that allow them to reach a new security level. Also, HAPI’s Oraclizing and DAO system delivers SaaS in the DeFi environment that prevents hack attempts.

Cybersecurity Triad and Cryptocurrency

Cybersecurity is a multifaceted field that encompasses a wide array of elements and transforms depending on the immediate utilization. The trifecta of cybersecurity however remains constant throughout digital space and directs the general axis of implementation. Confidentiality, Integrity, and Availability are the cornerstones of cybersecurity[1]. Normally, those three represent the main pillars cybersecurity is aimed at retaining, securing, and providing to respectively. In essence, if we were to describe cybersecurity by the current standards it propounds, it would be defined as follows: preclusion of adversarial digital attacks on private and governmental entities directed at acquiring confidential data or maliciously inhibiting the operational state of some elements within a given enterprise, institution, or entity.

The most common transgressions frequently include a hacking intrusion, leakage of data, and exploitability of a certain element in the system. The methods to intrude in the system or gain access to confidential data are multitude. Ranging from phishing, trojan, ransomware, and even social engineering that deserves a separate prelude of its own. These manifolds of potential attacks endanger users across multiple planes of internet usage. The protection, therefore, has become a prominent goal for everyone irrespective of the scope of business or individual preferences - cybersecurity is the most important field to be aware of in the current day and age.

One of the most novel and vulnerable targets for cyberattacks is the cryptocurrency market[2]. The reason for the onslaught of adversarial attacks on this market is overt - the novelty of the market and its accessibility inherently attract a large cohort of users who are uninitiated in the norms and caveats of this market. An unregulated

[1] *Perrin, Chad. "The CIA Triad." TechRepublic, TechRepublic, 30 June 2008, www.techrepublic.com/blog/it-security/the-cia-triad/.*

[2] *"The Chainalysis 2021 Crypto Crime Report." Chainalysis, go.chainalysis.com/2021-Crypto-Crime-Report.html.*

milieu also encourages the propagation of malicious actors and unpunished acts do exacerbate this propagation even further. Therefore the remedy of the issue looks twofold: firstly, there is a need to raise awareness about the cybersecurity aspect in the crypto sphere that effectively will “gamify” or engage players within the system to be involved in establishing security and safety by directly contributing to it. Secondly, the current notion of regulations is taken with a grain of salt by the crypto community as being invasive, denigrating to the ideals of privacy, and ruinous to the confidentiality of self-governed financing thus in order to upend this notion we need the implementation of a decentralized equivalent that will be impartial and non-custodian in its operation.

In order to gain a better insight into the necessity of such a system or protocol in place, we will examine and investigate the current state of crypto and reference it to the real world factors as well as the most notable historical occurrences that negatively affect and to some extent hinder adoption of the digital asset medium as a financial instrument of wide institutional and commercial adoption.

Overview

AML in and outside of crypto

The imperative and crucial notion of cybersecurity hinges on compliance-ready solutions. CEX incorporates a slew of preventative measures in hope of keeping exchange regulations compliant. One of the most common approaches towards eliminating any underhanded schemes is KYC. KYC creates an impediment for money launderers by instituting an identity disclosure that allows for easily traceable detection of potentially fraudulent activity. This however can be considered a "passively" constructed attempt at preserving compliance rather than proactively contending with the problem itself. It can be reasoned from the fact that AMLD5[3] (Anti Money Laundering Directive 5) is inundated with a plethora of blind spots that can quite handily be exploited either by the design of its inherently flawed structure or insufficient collaborative endeavors from the local and international regulatory bodies.

The vastness of the issue exacerbates further if we are to consider that the backbone of the KYC is private data aggregation that impedes the decentralized undertone of the cryptocurrency market. This coupled with mediocre security and constant leakage of private data is the scourge of the privacy conformant and conscious individuals.[4] The conclusion made by Houben and Snyer: “[...] this approach is not very convincing if the legislator is truly serious about unveiling the anonymity of cryptocurrency users to make the combat against money laundering, terrorist financing and tax evasion more effective”. In order to preserve the solidity of the non-custodial nature of the digital currency market, there is an increasingly

[3] *“EU's AMLD5 Is Not Enough.” Bruc Bond, 3 Jan. 2020, www.brucbond.com/article/eus-amld5-is-not-enough.*

[4] *Carrillo, Casey. “‘Probably the Largest Kyc Data Leak in History’ Demonstrates the Importance of Bitcoin Privacy.” Bitcoin Magazine: Bitcoin News, Articles, Charts, and Guides, Bitcoin Magazine: Bitcoin News, Articles, Charts, and Guides, 29 Mar. 2021, bitcoinmagazine.com/business/probably-the-largest-kyc-data-leak-in-history-demonstrates-the-importance-of-bitcoin-privacy.*

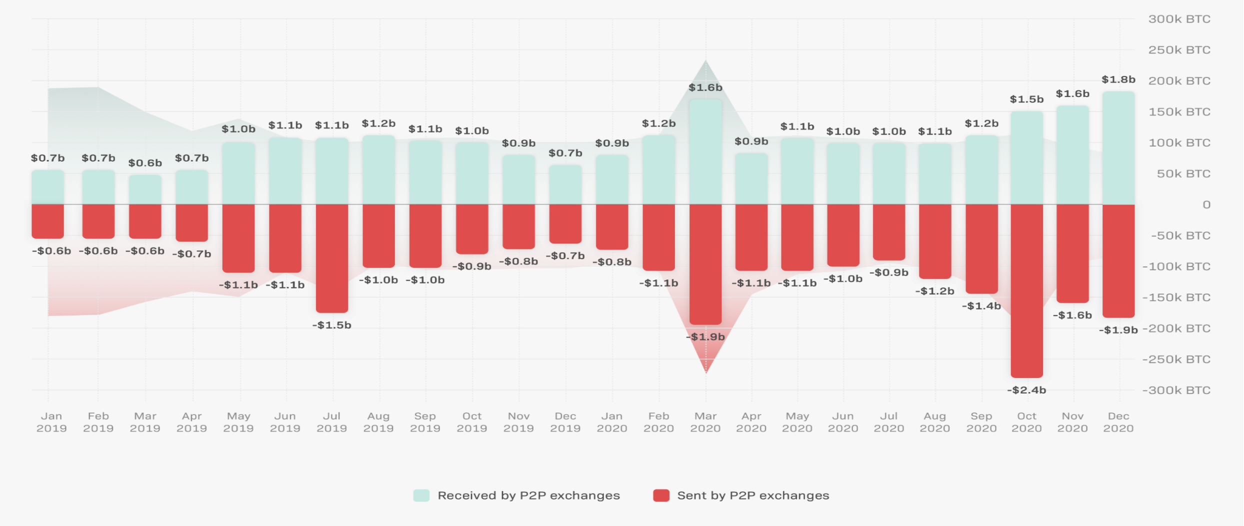
salient demand for, if not total substitution of the currently imposed and nascent suggested measures but at the very least a supplementary decentralized tool that would aid in promoting fair and compliant environment. With the emergence of the new type of financial instruments and subsequent adoption of these instruments by a wider cohort of people, there is a patent need for revisitation and reassessment of the customarily enacted compliance mechanisms and their applicability to the new paradigm of financing and investing. This is in part what Edward Kane means by referring to regulatory dialectics[5].

Despite CEX implementing KYC as a standard in order to enact at least in some domains a failsafe for money laundering[6], it only slovenly manages to do so taking under purview the importance of avoiding rigorous restrictions more so than the actual “secureness” of the platform. In the contemporary digital currency market and its rendition, a very implicative concern can be made about the uncertainty and general untrustworthiness of the digital currency equivalents. In the US alone there is a stark and relentless rise in malfeasant activities borne out from the crypto space. Specifically, new project launches have become a hotbed for gullible retail investors who are seeking a lucrative investment opportunity and ultimately fall prey to the derisively bold promises of enrichment of one's bank account. This can be attributed to the relative ease with which a potential investor can interact with the digital currency market circumventing centralized exchange and going straight to DEX, ill-equipped to the knowledge needed in order to be cognizant of incongruities with specifics of crypto investing. Relevantly, the outflow of retail investors from the centralized exchanges has also impelled a massive degree of popularity relay into decentralized alternatives of exchanges. In essence, DEX is poised beyond the reach of regulations and therefore is, on one hand, an unbridled way to trade and invest in digital assets but on the other hand is bristling with the unregulated in- and outflow of frequently fraudulent funds turnover. Unregulated and stripped of enforced compliance, DEX has garnered an unprecedented degree

[5] *Eisenbeis, Robert. “TNB and the Regulatory Dialectic.” Accelerating Progress, 14 Dec. 2018, www.spglobal.com/marketintelligence/en/news-insights/trending/bh8zuqyjenucv2z_azcaza2.*

[6] *“Are Centre Centralized Cryptocurrency Regulations the Answer? Are Cryptocurrecny Regulations the Answer? Three Countries; Three Different Directions .” Brooklynworks, brooklynworks.brooklaw.edu/cgi/viewcontent.cgi?article=1960&context=bjil.*

of recognition from market participants. This is attributed to the warranted fear about the unrestrained control and potentially oppressive influence of the governmental bodies, manipulating the to and fro movement of funds, completely omitting the individual jurisdiction over personal financial means. In this vein, DEX “entrusts" a complete power over the financial means of every individual to the very same individuals without constricting the medium of financing and investing with custodian measures. There is no clear statistical data on DEX being a “Haven” for money laundering as it is often stigmatized by the media, there is however a great chart on the volumes of BTC transactions on so-called P2P exchanges or DEX.

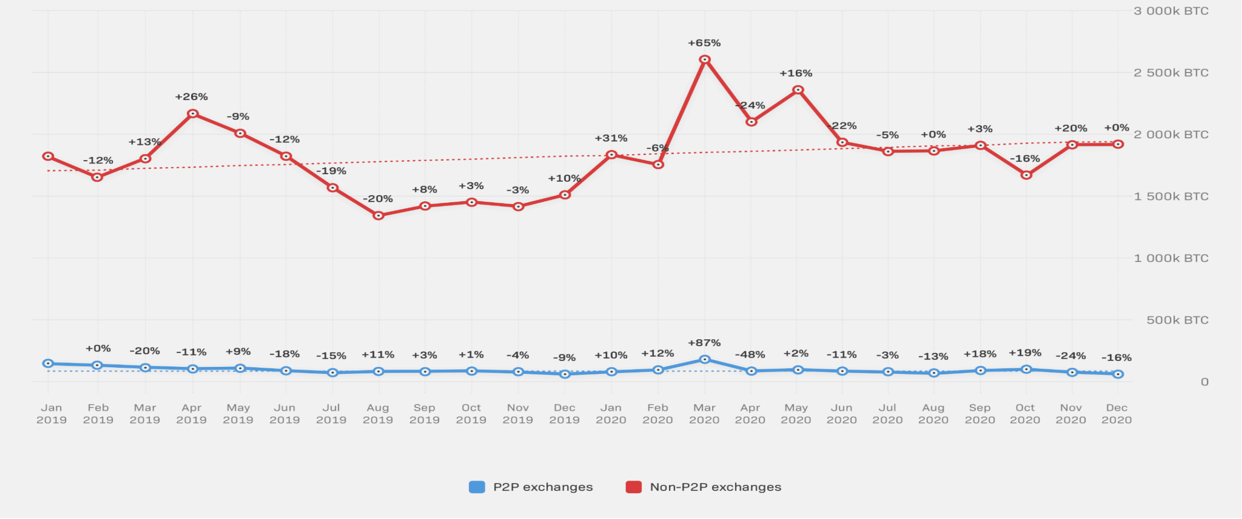


Source: Crystalblockchain.com

Crystal[7] makes a correlative derivation from the data claiming that with the FATF stating that they will take rein over the P2P exchanges in June 2021, there is a steep decline in received amounts of BTC overall compared to centralized exchanges, denoting a potential disinclination from money laundering. The recent Plenary of FATF long-term review of the implementation of FATF’s revised Standards on virtual assets, noting that there is clear progress to a more collaborative effort toward rectifying the immediate issue with potential ML and

[7] *“12 Month Review Revised FATF Standards Virtual Assets.” Second-12-Month-Review-Revised-FATF-Standards-Virtual-Assets-VASPS, www.fatf-gafi.org/media/fatf/documents/recommendations/Second-12-Month-Review-Revised-FATF-Standards-Virtual-Assets-VASPS.pdf.*

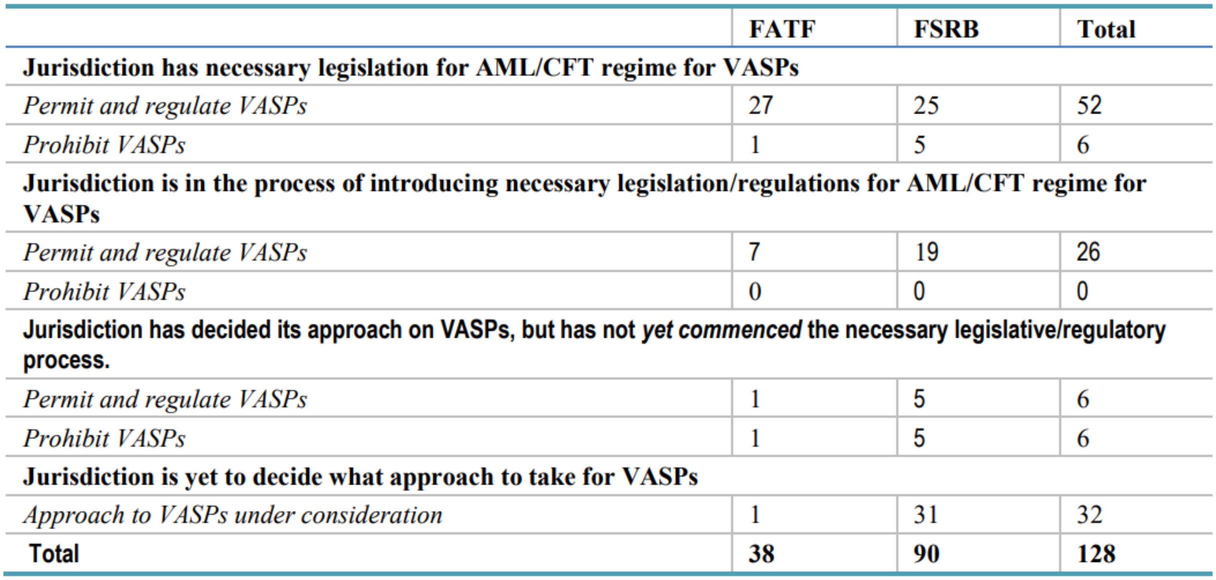
FT incursion. [...] “58 out of 128 reporting jurisdictions advised that they have now implemented the revised FATF Standards, with 52 of these regulating VASPs and six of these prohibiting the operation of VASPs. The private sector has made progress in developing technological solutions to enable the implementation of the ‘travel rule’.” Despite the progress being made, there is a predominantly listless reaction toward implementation of radical changes especially with the proposed “travel rule” that postulates a dangerous notion of “...all digital fund transfers creators and beneficiaries should exchange descriptive information” that unapologetically undermines the entire concept of decentralized exchange system and unavoidably creating an overt rebuttal from the DeFi community circles.



Source: Crystalblockchain.com

This unresolved issue creates a cleft between two groups namely market participants and regulators hindering mass adoption of digital ungoverned financing. While the goal of erecting compliant-ready solutions is valiant, the means proposed don’t align either with the community or the vision of the majority of projects themselves. Implementation of supervisory elements, centralized funds tracing heuristics, are derelicts of obsolete governmental overseers trying to impose strict and unnecessary bounds on the financial liberty of individuals. Despite FATF’s persistence on the introduction of custodian compliant protocols by propounding that potential future investors might be: “ [...} disincentivize further investment in the necessary technology solutions and compliance infrastructure”, there is neither proof nor contingency to this statement. Certainly, it is crucial to minimize the frequency of fraudulent, illegal, and suspicious intrusions in the crypto space. However, it necessitates finding an appropriate compromise that is reached inferentially instead of carelessly exposing the whole structural integrity of blockchain to the unneeded and redundant custody. It should, however, be noted that despite the constant demonization of any type of regulatory measures in the crypto community, the majority of regulators and NGOs are in fact against outright discarding P2P exchanges and instead innovate in a way that would ultimately be beneficial in the wide scope and prolific adoption phase.

Table 1. Progress in implementing AML/CFT regulatory regimes for VASPs[4]

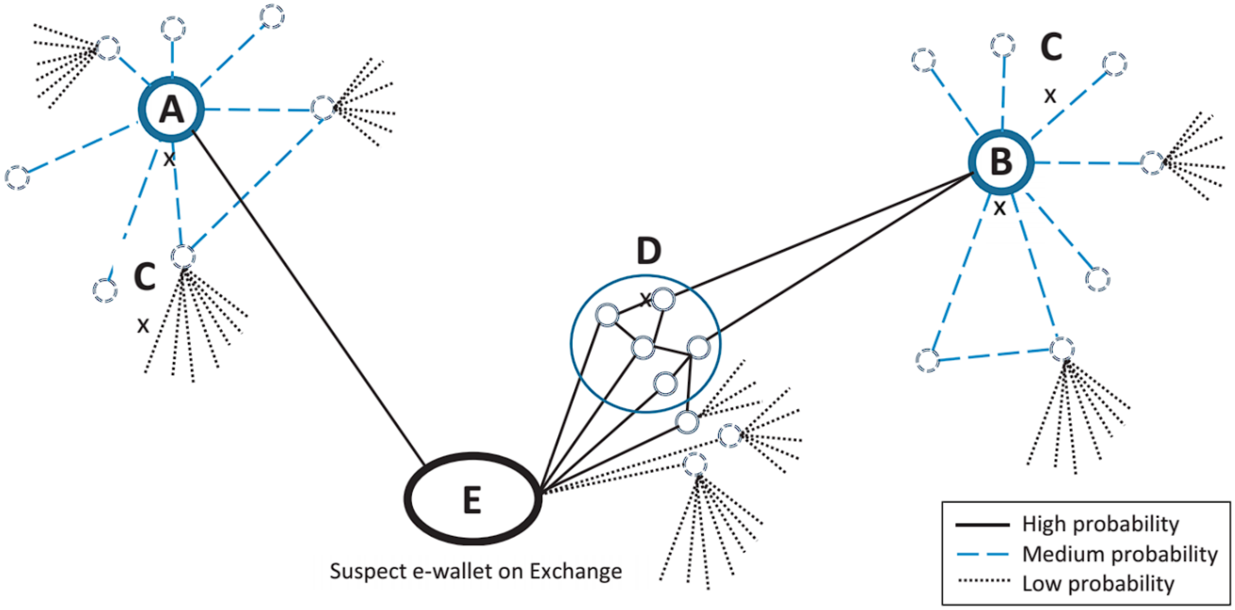


	FATF	FSRB	Total
Jurisdiction has necessary legislation for AML/CFT regime for VASPs			
Permit and regulate VASPs	27	25	52
Prohibit VASPs	1	5	6
Jurisdiction is in the process of introducing necessary legislation/regulations for AML/CFT regime for VASPs			
Permit and regulate VASPs	7	19	26
Prohibit VASPs	0	0	0
Jurisdiction has decided its approach on VASPs, but has not *yet commenced* the necessary legislative/regulatory process.			
Permit and regulate VASPs	1	5	6
Prohibit VASPs	1	5	6
Jurisdiction is yet to decide what approach to take for VASPs			
Approach to VASPs under consideration	1	31	32
Total	**38**	**90**	**128**

Instead of FATF attempts to standardize the priority of AML for VASP (virtual asset providers), there is a patent unwillingness and pushback to do so. Blockchain data analysis groups such as Ciphertrace and Chainalysis do have their foot in the door in the provision of data, transactions retracement, and even deanonymization of peer-to-peer interrelations. However, the main contentious point about them is a centralized aspect of the operation and post-factum interception/analysis.

In AML compliant exchanges all traders are subjected to the disclosed environment that is easily traceable. Suspicious transactional activity can be easily impeded if the necessary suspicions are raised either from the governmental bodies or detected by automated systems. Each E-wallet on CEX that conducts fiat-to-crypto trading is marked and therefore CEX has a wide reach in terms of deriving data on the actions of users within it. For instance, if there is a node under suspicion of laundering money on the centralized exchange the system can automatically flag it accordingly and enact a proper verdict on the spot. Subordinated nodal wallets indicate a lower risk and exude less suspicion in general but still remain under the watchful eye of the system. Additionally, there is a possibility for CEX to also inhibit structuring or smurfing attempts i.e.prevent dissemination or regroup of funds inside the exchange to different wallets becoming the focal point of the investigative forces of the system.

CEX also widely utilizes the above-mentioned layers of AML security solutions that are mainly directed towards assessing the faultiness and inconsistencies of digital transactions. CypherTrace in particular is able to actively infer the likelihood of illegitimacy. This is done with the help of a plethora of advanced detection algorithms that use proprietary clustering techniques to link a wallet to transfers. Chainalysis on the other hand exclusively focuses on tracking Bitcoin transactional activity by tracking and collating in and output transactions inside the ledger establishing chronological and historical order of all of the transactions that a particular address has been engaged in.



Tracking within Blockchain and Distributed Ledger Technology

There is an erroneous assumption that Blockchain, specifically Bitcoin, transactions are beyond the reach of tracking and deanonymization[8]. This, however, has been demystified by a broad range of papers that go into details about how these transactions can be identified and even the identity of users potentially being surfaced. Incidentally, even so-called privacy tokens can to some extent be traced back and subjected to the same fate as Bitcoin.

Dash[9] is one of the most well-known privacy-related forks of Bitcoin that claims to have a robust unidentifiable transaction path. In simplified terms, there is a combination of transactions when each sender issues one. By creating a single transaction and obfuscating each sender in particular, Dash claims to have a solid foundation of privacy and an identity concealment solution. Recent research examining the validity of privacy of Dash has been conducted, showing that the hidden recipient address can in fact be disentangled and such links can be created on the basis of the value being transacted.

Monero[10]'s intricate use of mixing can in fact obfuscate the destination of the transaction. One of the usages of Monero then is to muddle the retroactive path of the transaction and make it impossible to follow the lead. Nevertheless, Monero also suffers from potential deanonymization and users disregarding mix-ins or exploiting inferences about the age of the coins used as mix-ins. Zcash also represents one of the mainstays of privacy coins. Zcash uses a shielded pool hiding in the transaction the values and addresses of senders and recipients. It is possible

[8] *Tracing Transactions across Cryptocurrency Ledgers - Usenix. www.usenix.org/system/files/sec19-yousaf_0.pdf.*

[9] *E. Ben-Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza. Zerocash: Decentralized anonymous payments from bitcoin. In 2014 IEEE Symposium on Security and Privacy, pages 459–474, Berkeley, CA, USA, May 18–21, 2014. IEEE Computer Society Press | S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage. A fistful of bitcoins: characterizing payments among men with no names. In Proceedings of the 2013 Internet Measurement Conference, pages 127–140. ACM, 2013*

[10] *Hinteregger, Abraham, and Bernhard Haslhofer. "Short Paper: An Empirical Analysis of Monero CROSS-CHAIN TRACEABILITY." Financial Cryptography and Data Security, 2019, pp. 150–157., doi:10.1007/978-3-030-32101-7_10.*

though to if not completely strip the anonymity layer but substantially reduce its efficacy by identifying links between concealed and partly hidden transactions.

The objective of this paper, however, is not to debase the privacy value of cryptocurrency but instead showcase the readability of blockchain solutions that would enable a decentralized equivalent of security and compliance to be enacted. One of the actuators for this lies deep in the design of blockchain - distributed ledger.

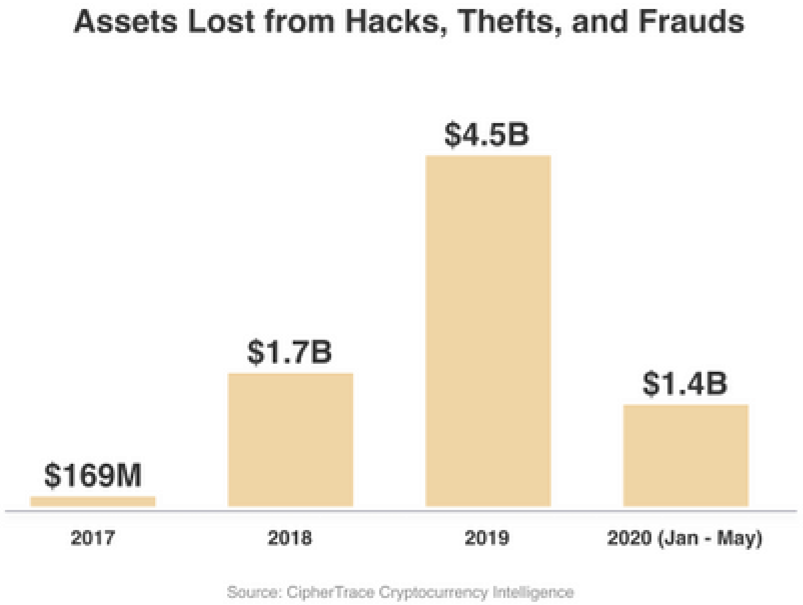
Distributed ledgers have allowed an unprecedented intersection of publicly available messages that coincidentally safeguard private and sensitive data such as identity, geolocation, and destination of the transaction. It is a permissionless, ungoverned, and fully self-sufficient storage that subsists on the activity within it. This embedded notion of openness paradoxically becomes the guarantor of the veracity of the whole system and essentially is the core element of blockchain technology. Unlike the centralized renditions, DLT is impossible to tamper with by the nature of the peer-to-peer design of the technology in question.

Public Ledgers made it possible to unite transparency and security while shifting the base framework of the process behind the scenes, eliminating intermediaries, and in some cases even cheapening the cost of an operation. Despite the robustness of the distributed ledger technology and its irrefutable quality of security and invulnerability to outside manipulation, there is still a myriad of methods utilized that are employed in order to undermine the rigidity aspect, not of the entire technology but rather individuals actively interacting within it. Specifically, Blockchain being codified and secured by the cryptographic hash functions is impenetrable to simple brute hacking attempts that are focused on deciphering it. Bypass approaches on the other hand can be manifold. Ranging from social engineering, wallet fishing, and negligence, to more sophisticated exploits via smart contract backdoor techniques. In this vein when talking about the security of the network it is essential to consider the outer layer of confidentiality, soundness of code, and absence of malevolently hidden loopholes or backdoors.

“Traffic analysis” or tracing of the activity on the public ledger of an individual wallet address can be leveraged effectively with the help of innate openness of transactional activity without impinging on the ironclad ideals of decentralized infrastructure and abiding by the privacy conformity of each player in the system. In the traditional cyber security field traffic analysis is the first stepping stone employed that enables and lends a fairly accurate representation of chronology and history of transactional intersections of a particular player. Logically, this basic method can seamlessly be translated into the blockchain realm and even arguably instated much more relevantly because of the above-mentioned "publicness" of the ledger.

Hacking Incursions on DApps and Liquidity

Hacking can be denoted as an act of unauthorized access to privately-owned or identity-confidential data. Traditional hacking outside the cryptocurrency market has gained traction by the sheer notoriety and glamourization of the act itself. From that stemmed a plethora of movements aimed at preventing and consolidating the security facet of the related tech and/or application. It can be argued that the issue with cybersecurity in cryptocurrency has started to surface very recently despite the galore of hacks that started to sprout from early 2013.



Source: CipherTrace Cryptocurrency Intelligence

[11]

Because of the relative ease of access and structurally different mechanism of work, the cryptocurrency market has become a vulnerable target for a range of sweeping attacks. Frequently[12], the main victims of such attacks are exchanges, liquidity providers or aggregators, and DeFi projects. These hacking intrusions have been growing exponentially in the last 5 years because of the substantial expansion of the cryptocurrency reach and the common ones include ransomware

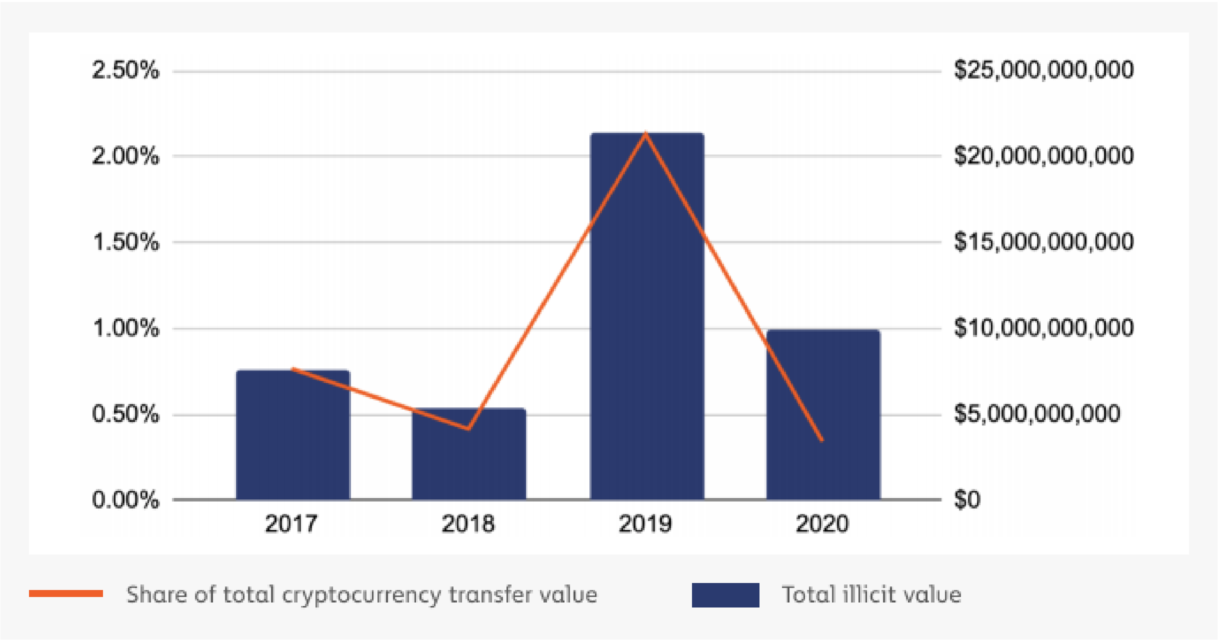
[11] Everything Financial Institutions Need to Know About Cryptocurrency Definitions, Obligations and Compliance Best Practices CipherTrace Cryptocurrency Intelligence June 2021

[12] *"Crypto Crime." Crypto Head, 5 July 2021, cryptohead.io/research/crypto-crime/.*

[13], scamming activity, phishing scams, and hacking of exchanges or wallets (Chainalysis 2019). The main vector[14] of attacks can frequently be delineated in either internal flaws of the code or negligence of shared private keys. Indeed, the amount of cybercrime involving cryptocurrencies has grown via ransomware, scamming activity, phishing scams, and hacking of exchanges or wallets.

Exchange hacks are one of the most financially devastating types of hacks and traditionally incur losses over a wide field of parties involved: immediate participants and entrants of an exchange, reputational loss of exchange itself, and distrust of the cryptocurrency market in general. The available data is clear that the egregiousness of the attacks can get fairly out of hand. According to Chainalysis, More than 20b $ were illicitly gained access to or interacted within 2019 and 10b in 2020.

Total cryptocurrency value sent and received by illicit entities vs. Illicit share of all cryptocurrency activity | 2020



[13] Huang, Danny Yuxing, et al. *"Tracking Ransomware End-to-End." 2018 IEEE Symposium on Security and Privacy (SP), 2018, doi:10.1109/sp.2018.00047.*

[14] *Huang, Sherena. "Cryptocurrency and Crime." FinTech, Artificial Intelligence and the Law, 2021, pp. 125–143., doi:10.4324/9781003020998-11.*

Normally, exchanges are shielded by a certain layer of protection that in most cases represents a typical tracing software or protocol that acts as an observer of fund flow. It can act on the risk posed by freezing a transaction and/or assess the individual risk of every incoming transaction by the set of proprietary algorithmic and dataset parameters.

In the case of unforeseeable events, there can be either a governmental involvement or a privately-owned cybersecurity investigation to trace suspicious fund transfers. The following happens when there is a suspicion about the hacking address involved:

1. The tracing starts from the known address
2. Following the address through up to a myriad of different addresses
3. Identifying the destination of the last address to hit an off-ramp service
4. Uncovering the identity of the perpetrator by issuing a subpoena to an above-mentioned service.

The problem with the chain of events above is the post-factum nature of it. Tracing can be beneficial in the spectrum of enacting enforcement of a law or in other words indictment of an individual behind the attack. Thus deanonymization becomes the core factor of the process. This arguably should not be the forerunning goal of the incident. Instead, there should be both preventative and reactive mechanisms in place in order to act on the attack immediately after it transpires or, in the best-case scenario, prevent it altogether.

There is a need then for a failsafe mechanism within the decentralized space that can vicariously or hand in hand execute a similar or supplementary role to the currently established centralized system of custody while eliminating the need to employ tight surveillance and deanonymization.

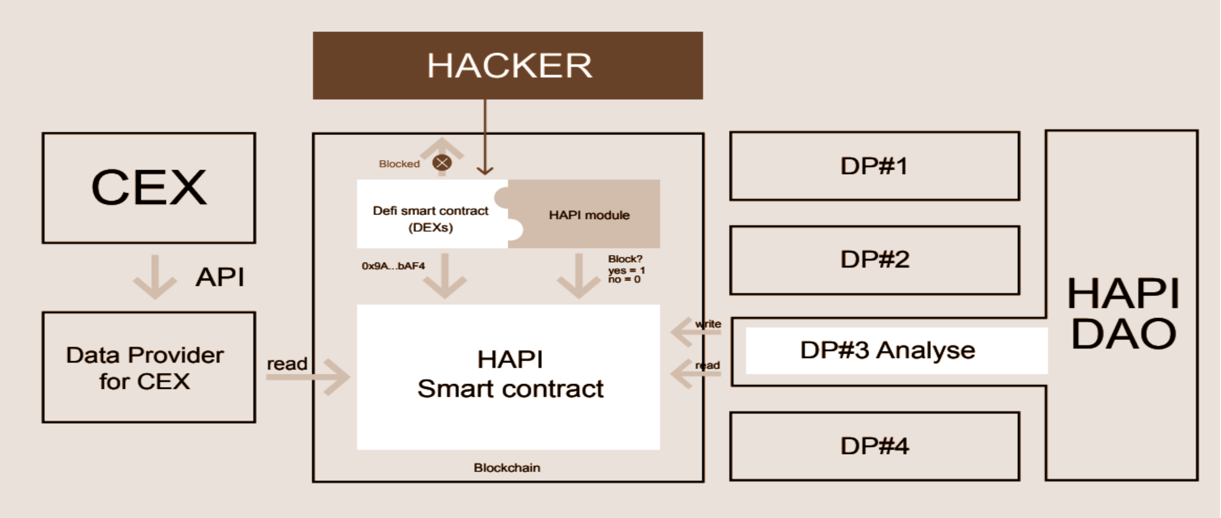
This can be fairly easily leveraged with the help of smart contract integration that records the inflow and outflow of addresses interacting within a given infrastructure. Each recorded address then can be granularly investigated either by an automatically embedded algorithmic system that employs various clustering techniques and factorial values or/and manually by a specifically designated group of individuals via a governance system. Various datasets are subsequently utilized to group and cluster transactions in different categories. The most basic categories might include dualistic representation of the nature of the transaction in the given time, namely “risky” and “non-risky”. The risky transaction is defined on the basis of previous activity of a given address and confluence of factors or datasets that contribute to that. For instance, a reputation scoring mechanism or RSP can be instituted in order to effectively and efficiently profile each address and endow it with the pertinent score. In this vein, “risky” transactions are considered risky because of a plethora of factors contributing negatively to the scoring.

Distinguishing a risk factor requires a meticulously crafted pattern recognition algorithm. This is crucial as to not misnomer a specific address that might engender a chain of unfortunate misconstructions. In order to achieve granularity, specificity, and accuracy to assign an address with, there is a need to query the history of past transactional activity. Concurrently, we need to find similarities and appropriately compare them to glean any results. Whereupon we can identify a propensity factor of the given address out of similarities found and infer that this address might in fact be a “risky” one. This is also a juncture where machine learning techniques might be handily used. Machine learning will significantly simplify the manual work needed for dataset provision and augment the future detection capabilities of the system. Transactions are analyzed based on the preemptively inputted set of quantifiable characteristics that are consequently also subclassified with the help of metrics. Machine learning will not only append to the dataset but also strengthen the accuracy of classification meaning a more robust automated system can be attained.

Per regulatory dialectics, there is a need to constantly revise, add or refine certain parameters in order to be congruent with unavoidable alterations. These alterations can include administrative decisions, compliance regulations, or even locally enacted laws. In this vein, a system is able to be fully customizable and tweaked on the fly to adapt to the regulatory instances of change. It is also mandatory to add a kind of DAO system in order to make the system decentralized, incentivized by communal decisions of the majority, or in other words, democratized, and malleable enough as to enable an ever-growing, self-subsisting body of decentralized autonomous regulations to be implemented and defined by community voting.

HAPI - Core of Decentralized Cybersecurity

HAPI is an autonomous, and decentralized standard that assumes the role of ungoverned equivalent of centralized solutions that are absent from the contemporary blockchain sphere. Specifically targeting the most controversial moot points and comments on blockchain technologies and decentralized applications lacking in reliant measures against money laundering, high risk-exposing investment in unreliable ventures, and general apprehension around the vulnerability of the applications built on various blockchains. The cascading expansion of the digital currency market, as well as incessant growth of adoption in the coming years, evince the need for a non-governmental but trustworthy institution of analogous security solution that in the event of unexpected malicious activity is able to actively react and mitigate, alleviate or completely preclude from happening the loss of funds.



The solution HAPI presents is of paramount importance for two reasons: lack of openly distributed SaaS security measures in crypto, and decentralized community-dependent voting system that defines the axis of governance, and provisions of the database. The database is crucial for HAPI protocol and its long-term growth because it will allow HAPI to be constantly front-running fraudulent activity and notify those in the system about potential dangers. Anyone can submit an illicit address into the database thus contributing to the overall "secureness" of the space. This formula is further incentivized by rewarding participants with respective native tokens of the HAPI Protocol. Practically, HAPI emboldens and encourages even uninitiated but affected people to take part in preserving a certain level of security and cybersecurity-proofness of the digital asset medium. In essence, the "collaboration of everyone" will theoretically not only create a safer environment but also will become a tremendous catalyst of awareness about cybersecurity in general. This concept is buttressed by our vision of simplifying the way people report malfeasance and malicious activity. We aspire to design a straightforward and credence-backed consortium that will be open for those willing to partake and vote in the democracy-driven modus operandi.

We also realize the obstacles in our way to creating an efficacious cybersecurity protocol whilst retaining the decentralization and non-custodial nature intact. In order to achieve that, we need to abstain from directly exposing the identity of a perpetrator even if it means, in some regards, non-compliance to regulations. The glaring issue of the centralized attempts at apprehending and halting unsanctioned or malicious fund turnover is the palpable focus on deanonymization. On the basis of deanonymization being a critical detriment to decentralization, it would be morally reproachable for a project built on these tenets to disregard them. Therefore, HAPI aims at retaining these crucial components and staying truthful to the anonymity and identity insulation by solely incurring negative consequences on the addresses and not individuals. Deanonymization is not taken under purview by HAPI Protocol therefore we do not intend to intrude into the private sector of individuals or demarcate specificities of use. Each network participant of the HAPI protocol is at the liberty of their own deliberations. HAPI Protocol itself doesn't segregate, which in essence makes it publicly and freely distributable and accessible, into groups neither the usage of the protocol nor the users. Since HAPI Protocol is first and foremost focused on the B2B (Business to Business) area of influence, we are merely a provider of services that can be utilized indifferently to the betterment of the aforementioned "user" or "business". HAPI Protocol is autonomous and is not a regulatory body that can enforce or impose any restrictions by itself but instead is driven by a voting system of governance.

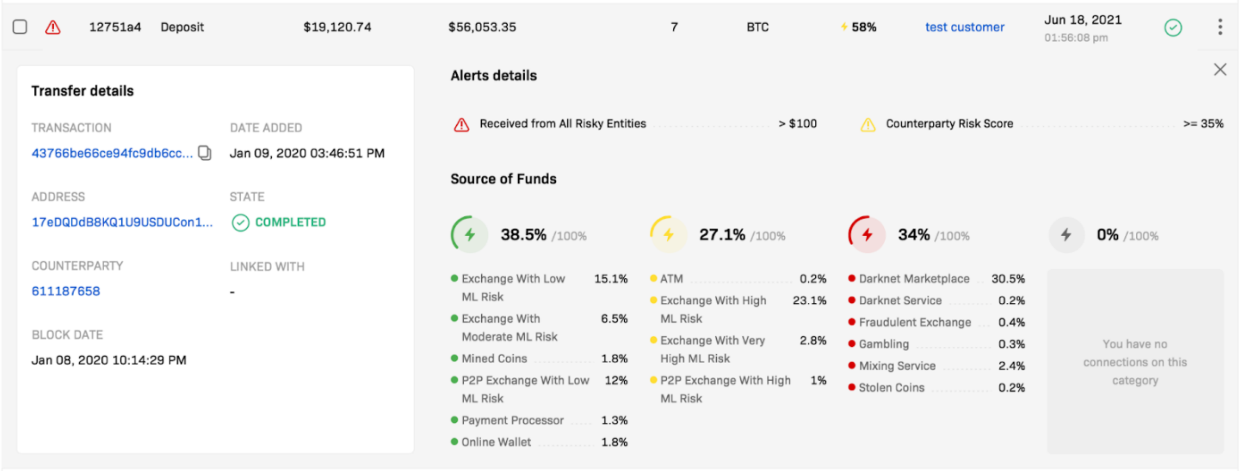
Because of the essence of HAPI focusing on the B2B sector and, if HAPI's use is to become extensive enough, it can be reasonably assumed that the frequency and severity of hacks can be significantly curtailed because of the less or no options to withdraw illicitly controlled funds.

HAPI will also be a crucial aggregator of illicit addresses and publicly distribute them across a number of CEXes and DEXes in a timely manner. This will help DEX and CEX to blacklist certain addresses from laundering money on the platform. CEX and DEX will automatically receive notifications about the address that has been flagged as risky which will create an overarching, unified system that will ensure automated up-to-dateness on fraudulence across all exchanges.

Data Providers

The essential quality of HAPI is database building. In order to build a robust database that will reliably supply smart contracts with verifiable and constant provision of addresses, there is a need to aggregate this data. There is however an issue relative to the time needed for build-up. This process of acquiring data is a long-term one therefore it hinders us from instituting our own database from the outset. As such we are resorting to the use of a third-party data provider that will be a provisional substitute for the time of our own data aggregation. Data provider therefore can be described as a centralized tracing solution that will provide up-to-date data on the blockchain activity and relay this data via oracles to the HAPI SC (smart contracts). By this formula, we can largely circumvent the need for preliminary aggregation before deployment of the protocol and allow it to gather data in a, so to speak, real-life environment.

A data provider is a complex tool that enables a multi-granular analysis of blockchain transactions in real-time by utilizing a slew of intelligence solutions. The core element of the software is automation. It uses automated tracking that initializes tracing of the flow of digital funds across a wide field of intermediate wallets and pinpointing the end-point wallet. It also is able to effectively parse the blockchain into certain datasets that can ultimately be used as a foothold for identifying the likelihood of illicitness. The solution specifies a target, a certain in- and outflow of funds from the specific starting points, for instance, hacked or exploited DeFi project and endpoint, or also known as off-ramp, CEX, or any other entity that enables a withdrawal of fiat.



□	⚠	12751a4	Deposit	$19,120.74	$56,053.35	7	BTC	⚡ 58%	test customer	Jun 18, 2021 01:56:08 pm	✓

Transfer details

TRANSACTION
43766be66ce94fc9db6cc...

DATE ADDED
Jan 09, 2020 03:46:51 PM

ADDRESS
17eDQDdB8KQ1U9USDUCon1...

STATE
COMPLETED

COUNTERPARTY
611187658

LINKED WITH
-

BLOCK DATE
Jan 08, 2020 10:14:29 PM

Alerts details

⚠ Received from All Risky Entities	> $100
⚠ Counterparty Risk Score	>= 35%

Source of Funds

38.5% /100%

● Exchange With Low ML Risk	15.1%
● Exchange With Moderate ML Risk	6.5%
● Mined Coins	1.8%
● P2P Exchange With Low ML Risk	12%
● Payment Processor	1.3%
● Online Wallet	1.8%

27.1% /100%

● ATM	0.2%
● Exchange With High ML Risk	23.1%
● Exchange With Very High ML Risk	2.8%
● P2P Exchange With High ML Risk	1%

34% /100%

● Darknet Marketplace	30.5%
● Darknet Service	0.2%
● Fraudulent Exchange	0.4%
● Gambling	0.3%
● Mixing Service	2.4%
● Stolen Coins	0.2%

0% /100%

You have no connections on this category

[15]

Our Data Provider is able to classify transactions based on a plethora of factors that ultimately play a role in appointing a respective verdict on their "safety". To that can also be added the ability of a data provider to ascertain the likelihood of the transaction being ML or not. Nominally every transaction is considered to be neutral or non-risky. Depending on the movement of the fund and transactional activity, a tentative verdict can be placed. Since a data provider is able to access (via openness of the ledger) each transaction path with relative ease, there is already accrued data on previous transactions. In this vein, a system is capable of employing historical and chronological methods of identifying the present likelihood of fraud. If a given address has been flagged or marked as "risky" a respective alert is issued, notifying HAPI SC about the potential insecureness or malfeasance of the address in question. Despite a swath of categories used by a data provider, HAPI will package all of the categories and reassort them into only two categories: “risky” or “non-risky”. More on categorization will be expounded in the later paragraphs.

To trace transactions we require an identificator of the deposit transaction on the input, normally called curIn (deposit) blockchain. To accurately identify an on-chain transaction we need two main conditions fulfilled.

[15] Crystal - https://crystalblockchain.com/investigations/how-crystal-investigations-truly-make-a-difference/

- The timestamp of a transaction is closer to the point of it being called-in via API
- The called-in value was the same as the value carried

An example of a request and response of a simple transaction tracer within Ethereum for one transaction thus will look similar to this:

```
Request: curl --data '{"method":"trace_call","params":[{ ...
},["trace"]],"id":1,"jsonrpc":"2.0"}' -H "Content-Type:
application/json" -X POST localhost:8545
```

```
Response: {
  "id": 1,
  "jsonrpc": "2.0",
  "result": {
    "output": "0x",
    "stateDiff": null,
    "trace": [{
      "action": { ... },
      "result": {
        "gasUsed": "0x0",
        "output": "0x"
      },
      "subtraces": 0,
      "traceAddress": [],
      "type": "call"
    }],
    "vmTrace": null
  }
}
```

In order to identify whether all inputs are related to the same entity in question, the multi-clustering method is often used as a determinator of all inputs belonging to the same end-point. The issue with this method, however, is that multiple clusters can be operated in each cryptocurrency. This complicates further if we are also to consider that each currency may also have a different “behavior” meaning that it cascades complexity even more making it harder to capture overarching data on the activity.

Therefore for us to accurately identify the relation between two or more addresses we need to find similarities. The easiest one would be to establish a common social relationship between transactions/addresses. For example, we shall assume that there is a recipient address that receives coins from two or more addresses in the curOut (withdrawal) or those addresses receive coins from the identical address in the curIn, we can reasonably assume that these addresses have in fact a common social relationship on the basis of having identical interactions within short timestamp. Although we denote the “relatedness” quality of addresses, it doesn’t mean however that we can clearly identify the meaning of each. There can be a number of potential meanings to this including two users simply sending coins to the address unknowing about each other's activities. Despite the meaning and goal of transactions remaining unknown, we can safely conclude that they do represent a kind of relatedness.

From this point on the analysis takes place utilizing proprietary algorithms and machine learning in order to sift, parse, detect, identify and categorize each transaction before and relay the results to the HAPI SC.

HAPI SC

HAPI embeds a set of smart contracts that in essence operate on the grounds of data provisioned to them via oracles. Generally, smart contracts employ a slew of predetermined lines of code that execute a string of functionalities and in the case of HAPI, one of the SC incumbent duties is to inform a project, CEX, or DEX about the likelihood of malevolence. HAPI SC operates adjacently to, for instance, DeFi smart contract. Every time a potentially hazardous or fraudulent transaction takes place, HAPI Protocol is able to detect it and write it in the SC, placing it in the “risky” category and preventing further unlawful activity from this address by essentially blacklisting and informing the network of HAPI Protocol linked entities that the address in question can, in fact, pose a substantial risk.

In a similar fashion, by extracting or querying data from various sources (including database provisioned via Data Providers), HAPI can effectively preclude a foreseeable danger from transpiring.

It should be noted that HAPI is several SCs working conjointly. Each SC is assigned a particular role. As we discussed above, one of the SC will essentially be the storage of data on addresses that also bears a grunt of categorization of the same addresses according to the nature of the risk they may pose. The basic set of smart contracts in HAPI Protocol:

- Core Smart Contract - address data storage, request/report handling, oracle interaction
- Token Smart Contract - utility token for core contract operation and governance
- Governance Smart Contract - endpoint for governance infrastructure for core contract management

Since we have already covered Core Smart Contract functionality to a sufficient degree and its main task of storing and efficiently classifying data according to the binary structure, it is reasonable to cover the remaining two SCs that are essential for the operation of HAPI Protocol.

Token Smart Contract is a standard asset tokenization smart contract that will be separately deployed on each blockchain since every Blockchain has its own a) Language for SC; b) Differences inherent to Blockchain's infrastructure (consensus algorithm, processing). The integral idea of HAPI however is to establish an interoperable, cross-chain ecosystem that operates blockchain-agnostically. Therefore the bridge functionality is of the utmost importance and will aid in instantiating robust interconnectedness between networks.

Governance Smart Contract essentially represents a governance-specific mechanism that will shoulder all of the related operations within it such as votes delegation, issuing proposals or submitting data in Core Smart Contract, and processing fees imposed when submitting the data.

Dual Categorization Mechanism in HAPI’s SC

The backdrop of classifying the transaction's likelihood of being fraudulent is denoted by two simple categorical subdivisions. A transaction can either be "risky" or "non-risky". The binary system of categorization enables an easier, more streamlined, and lower overhead of the processing system that will be sufficiently effective in denotation.

The long-term goal, however, is to construct a more complex system that will include a far wider domain for categories and will further subclassify them into case categories. This will include:

- Pending Assessment of transactions. In basic terms, it indicates that a given address is provisionally terminated from accessing an entity in question (CEX, DEX, or DeFi project, for instance, liquidity provider), for the time being, transaction and, particularly, an address assessed, is under meticulous investigation of the HAPI governance system and its members, commonly addressed as a committee.
- Tier-dependent categorization based on datasets.

Risk Assessment

Qualitative risk assessment plays a crucial role in the categorization of the likelihood of a certain adversarial attack. It also helps in mitigating low-risk exposures and avoiding a misnomer of an address. Therefore enabling a relatively transparent organizational method of transactions based on the risk they present is paramount. For that, we need a mechanism that would allow us to accurately, collectively, and categorically sort each potential adversarial attack based on the risk factor. Therefore we decided to introduce the Relative Scoring Mechanism (RSM).

With RSM each address is gauged based on the totality of transactional activities and the risk can exponentially increase or decrease depending on the "tier" of a risk they have engaged in. In essence, then, each transaction is appraised and given out a "score". With a certain threshold exceeded, the address can be assigned a "risky" index that would delineate that a given address can pose a substantial threat.

The risk score is calculated based on the number of transactions initiated by a user "T" and the quantity of those transactions that exceeded the threshold of risk represented by "r".

Simple Risk Assessment: $R = \frac{r}{T}$

An issue arises with new users who have non-existent scores because of the absence of any prior transactions. The limitation of this kind is also further aggravated if we are to consider that one non-threat transaction will immediately change the score of a given user to 1 which is the maximum possible value that renders a particular user "risky". The opposite also engenders a problem since one misgiving transaction on the side of a user will cause the score to substantially go down. That might create a discordance in assessment since a theoretical user with a large amount of non-risky transactions and, assuming, one risky will be deemed by the system as a more risky player than a user with only one non-risky transaction in the totality of his/her transaction history. Therefore the trust factor of our simplistic assessment mechanism might invalidly consider the former user as posing more threat than the latter despite the first user leaving more footprint and more room for qualitative assessment overall.

In order to rectify this issue, we need to discern and examine a transaction in a more complex and complete way.

Complex Risk Assessment with the introduction of Case Category (Reputation Score):

$$S = \begin{cases} 0, \ T = 0 \\ T * \left(1 - \frac{r}{T}\right), \ T > 0 \end{cases},$$

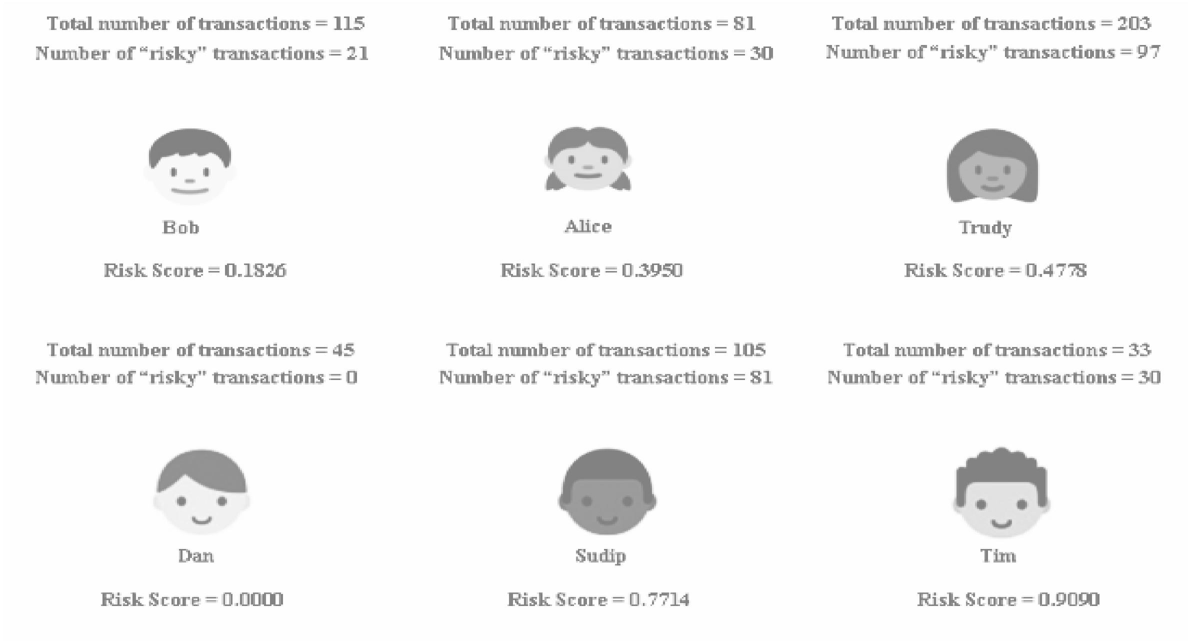
S = defines the reputation score of a given address. The reputation score is calculated by considering values such as the total amount of transactions by the same user and the percentage of those transactions being risky. The threshold of risk is calculated separately for each address based on the tiers and their respective points. The binary system of risk calculation still remains so there are only two possible outcomes: risky or non-risky. T = refers to the total transactions issued by the given address. R = pertains to the number of transactions issued by the address that involves any of the risk tiers and supersedes the threshold of risk.

Each tier within the Case Category is representative of and adds to the totality of score calculation. Based on the score and whether this score indicates a higher or lower risk exposure, the added “points” will influence the verdict of the assessment.

Case Category (Reputation Score)

A case is a reported event of illicit activity that encompasses an array of addresses. Each case is an instance of a recorded transaction of a given address in HAPI SC as well as its immediate transactional activity recorded on the blockchain.

Employing Reputation Score is salient in two ways: firstly, it empowers a more reliable gauge on the address' risk of posing a threat, and secondly, it works complementary to the simplistic structure, making less room for error.


Total number of transactions = 115
Number of "risky" transactions = 21
Bob
Risk Score = 0.1826
Total number of transactions = 81
Number of "risky" transactions = 30
Alice
Risk Score = 0.3950
Total number of transactions = 203
Number of "risky" transactions = 97
Trudy
Risk Score = 0.4778
Total number of transactions = 45
Number of "risky" transactions = 0
Dan
Risk Score = 0.0000
Total number of transactions = 105
Number of "risky" transactions = 81
Sudip
Risk Score = 0.7714
Total number of transactions = 33
Number of "risky" transactions = 30
Tim
Risk Score = 0.9090

Example of Risk Scoring

Exact points given and classification in general might be subject to change.

Tier 0 - No risk

- 0 Safe - this is a safe address

Tier 1 - Low risk

- 1 Wallet Service
- 2 Merchant Service
- 3 Mining Pool
- 4 Low Risk Exchange

Tier 2 - Medium risk

- 5 Medium Risk Exchange
- 6 DeFi
- 7 OTC Broker
- 8 ATM
- 9 Gambling

Tier 3 - High risk

- 10 Illicit Organization
- 11 Mixer
- 12 Darknet Service
- 13 Scam
- 14 Ransomware
- 15 Theft

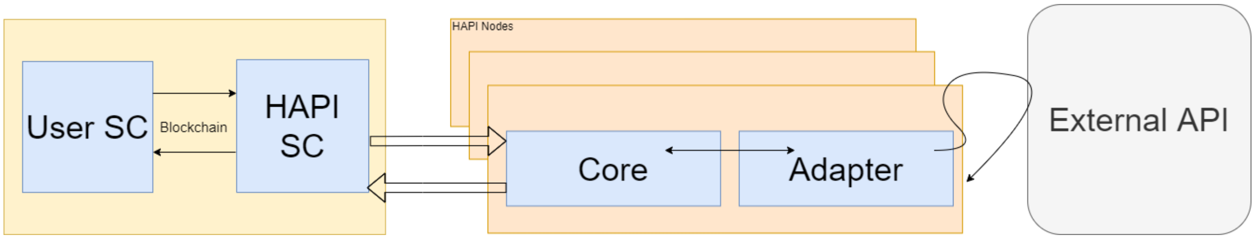
Tier 4 - Severe risk

- 16 Terrorist Financing
- 17 Sanctions
- 18 Child abuse

HAPI Oracles

Traditionally, oracles are the suppliers of off-chain data and tether the on-chain "world" with external or off-chain data feed. The main objective of oracles then is to reliably, consistently, and immutably provision data to the smart contract whilst keeping the decentralized notion intact. In the contemporary blockchain medium, oracles play a crucial role in bringing transformative innovation in the industry by linking a confined blockchain infrastructure with the price, state, and/or condition of the necessary events. The most frequent use of oracles being a supply of data on price feeds.

The operation of oracles is normally hinged on the simple linkage with data providers. If a certain DApp requires some kind of data that is unavailable on-chain for their use, they normally would contact an oracle provider that has an existing oracle network in place. In the case of oracles, a network refers to the collective number of node operators that uphold the operative state of oracles i.e. individuals running an executable on a personal computer to gather data and relay it to the blockchain. Oracles have their own reputational contract that informs an issuer and penalize a particular player if the data relayed has been tampered with in any way. The symbiotic relation between numerous node operators and the absence of one centralized issuer of data encapsulates the decentralized quality of oracle services. Data validation is also present in the oracle's structure. The validation of data normally happens on the aggregating contract whereupon the data is being assessed on the potential incongruities based on performance and behavior.

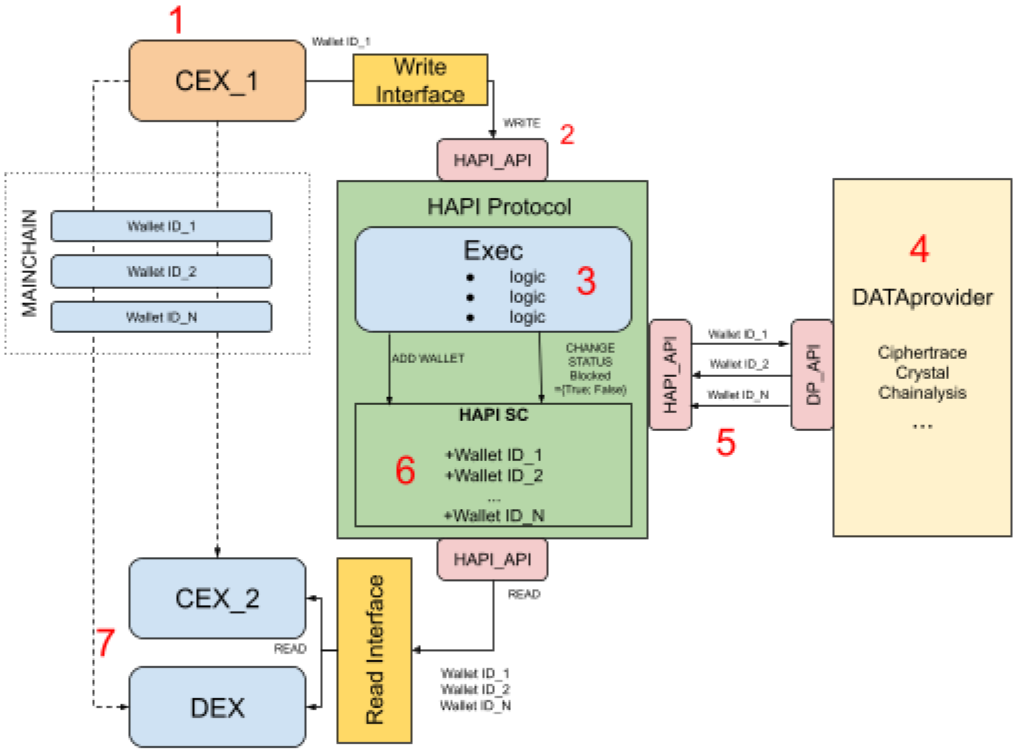


Within HAPI Protocol oracles are used to provide data from the data provider's existing database on the wallet addresses of the potential misaligned behavioral patterns. HAPI will use its proprietary oracles that will be fully decentralized and operated by the nodal system. Each node will represent a separate computer running an oracle service and provisioning off-chain data to the HAPI SC. Oracles providers, commonly known as node operators, will also be rewarded commensurable to the work done. The payment is issued in the native digital token of the protocol for the review and audit work done on the submitted data.

HAPI oracles are cross-chain and will operate with the most widely used blockchains. The main distinguishing factor of HAPI oracles is their utility. The main notion of oracles in the protocol is the provisioning of data to the smart contracts therefore unlike the accuracy and contingency nature of price oracles, the crucial aspect and quality encapsulated in HAPI-specific oracles is the speed with which data is provisioned.

Example of Attack on CEX and the Framework of HAPI Protocol

For the sake of imparting the clearest and the most visually understandable workflow of the protocol, it's reasonable to sequentially order the theoretical process of adversarial attack on CEX and DEX.



1. CEX 1 detects fraudulent transactions - hack that speedily turns into outflow of funds to the unknown wallet.
2. CEX_1 reports the event into HAPI Protocol that writes into the HAPI SC WalletID_1 to which the funds have been transferred.
3. HAPI Protocol analyses (based on the given metrics) a transaction and categorizes it in: Approved address WalletID_1 on the HAPI SC; Rejected inquiry of WalletID_1 on the HAPI SC
4. If the address is on the HAPI SC - HAPI Protocol sends WalletID_1 to DATAprovider in order to track and trace the upcoming transactional dealings from the same wallet.
5. DATAprovider via Oracles relays to the HAPI Protocol all of the nascent wallets on which the transactions are being transferred from the source wallet WalletID_1
6. HAPI Protocol analyses (based on metrics) wallet addresses Approved address WalletID_1 on the HAPI SC; Rejected inquiry of WalletID_1 on the HAPI SC
7. HAPI Protocol sends CEX_2 and DEX identificators of wallets from HAPI SC

An unlawful actor makes an attempt to send fraudulent funds on CEX 2. CEX 2 knowing the threat from the aforementioned wallet can expose it to the potential risk - blocks the transfer. Unlawful actor makes an attempt to connect the address to DEX, DEX declines this attempt and/or completely blocks the wallet from making a transaction within DEX.

HAPI allows for CEX to promptly notify its users about fraudulence taking place via Write Public Interface. CEX receives data about fraudulent activity described above, with this data available CEX may not only deftly freeze the siphon of funds but also reactively respond and announce to the users to abstain from using the CEX functionality for the time being. In this quite simple yet convenient fashion, CEX garners not only additional “points” for trustworthiness but also safeguards its users against the potentially exacerbating exploit.

Key Actors

Tracers - entities that trace on-chain data including transactional activity and wallet addresses. With ability to flag, notify, and mark certain patterns of potentially illicit on-chain activity. These notices are then assessed and a verdict placed in the span of 24 hours.

DAO Team - consists of committee and members of the HAPI governance organization. The DAO Team is responsible for relaying, assessing, proposing, and enacticting propositions based on a plethora of factors. In essence DAO Team represents the foundation of a community incentivized cybersecurity database that is aimed at providing a self-subsisting and ever-growing storage of malicious addresses.

Oracle - suppliers of the off-chain data. They play a crucial role in merging two separate entities, namely Blockchain and data outside the boundaries of the blockchain. The data oracles relay in HAPI Protocol is unilateral, and mainly concerned with querying data from data providers and sending it to Mainchain SC.

Data Provider - provisional measure that allows HAPI Protocol to work "out-of-the-box" instead of waiting for a database to be amassed. Generally, a data provider is a purveyor of databases on illicit addresses. They also take under purview analysis, classification, and parsing of the on-chain data.

Machine Learning

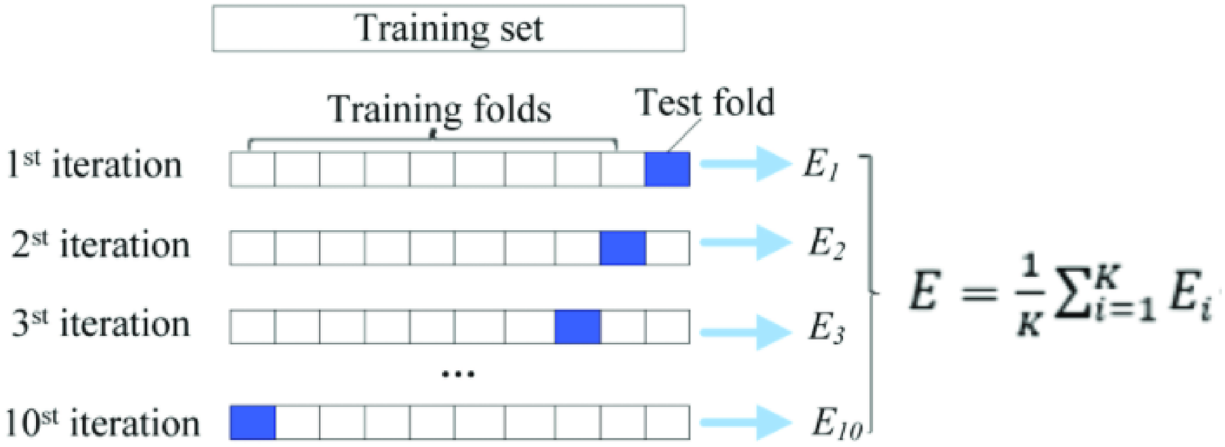
The utilization of machine learning algorithms is immensely useful in different aspects of data categorization. Categorization or classification algorithm is able to achieve exactly what HAPI Protocol needs - less computational overhead which also means increased data processing and higher specificity and accuracy of detection. At this juncture, pattern recognition is a crucial aspect of identifying the transactional history of an address, ramifications of the transaction paths, and commonalities in behavior. These similarities in behavior are specifically what we are interested in since they will, potentially, enable HAPI Protocol to operate at almost instantaneous speeds, allocating even more time for entities using HAPI to act on the illicit activity.

The complex issue arises when there is a need to utilize machine learning but there is not sufficient data on fraudulent transactions, for that reason, the first iteration of machine learning implementation will have a binary classification algorithm that will only have two potential predictors 0 or 1.
1 - identifies that the transaction in question is in fact fraudulent or poses a potential risk and 0 - refers to a safe risk score.

The core of the model then is created using customary variables, for the sake of simplicity and visualization, we will use 10 common variables: ***block_timestamp, block_n_txs, n_inputs, input_sum, output_sum, n_outputs, output_seq, and input_seq*****.** These will represent the base of the model. However, in order to refine and increase the accuracy of the baseline, we will need to employ a feature engineering technique that will help to achieve both granularities in classification and lay the foundation for model building.

Although the Supervised model of machine learning is ideal for our use since our model possesses a specific dataset, it is fraught with one issue in our case that may hinder accurate prediction - overfitting. The issue lies in the overdependence of the model on the very same predetermined datasets on which the model is being trained. To rectify overfitting issues effectively we need to apply a cross-validation workflow. In essence, cross-validation allows us to experiment on the independent dataset while holding out the test data enabling an effective partitioning of data. The data is being split into the following percentages: 80% for train purposes and 20% for test sets.

We use 10-Fold Cross-Validation and partition data into 10 separate parts.



In order to simplify the process of stratification of data and efficaciously partition it we also use StratifiedShuffleSplit.

Example: class `sklearn.model_selection.`**`StratifiedShuffleSplit`**(*n_splits=10*, ***, *test_size=80%*, *train_size=20%*, *random_state=None*)

Traditionally for classification model evaluation, confusion matrix is used. The basic principle of confusion matrix is to visualize the system confusing two classes. In our case, the confusion matrix can be appropriately used in order to ascertain whether the prediction mechanism is working as intended. Each instance in the confusion matrix is represented by two classes: actual and predicted. In this vein we can train the system to more accurately predict a particular address' likelihood of being fraudulent, in other words, increase specificity and consistency.

```
eval <- evaluate(d_binomial,
                 target_col = "target",
                 prediction_cols = "prediction",
                 type = "binomial")
```

```
eval
#> # A tibble: 1 x 19
#>   `Balanced Accuracy` Accuracy    F1 Sensitivity Specificity `Pos Pred Value`
#>                 <dbl>    <dbl> <dbl>       <dbl>       <dbl>            <dbl>
#> 1               0.551     0.58 0.672       0.632       0.469            0.717
#> # … with 13 more variables: Neg Pred Value <dbl>, AUC <dbl>, Lower CI <dbl>,
#> #   Upper CI <dbl>, Kappa <dbl>, MCC <dbl>, Detection Rate <dbl>,
#> #   Detection Prevalence <dbl>, Prevalence <dbl>, Predictions <list>,
#> #   ROC <named list>, Confusion Matrix <list>, Process <list>
```

```
conf_mat <- eval$`Confusion Matrix`[[1]]
conf_mat
#> # A tibble: 4 x 5
#>   Prediction Target Pos_0 Pos_1     N
#>   <chr>      <chr>  <chr> <chr> <int>
#> 1 0          0      TP    TN       15
#> 2 1          0      FN    FP       17
#> 3 0          1      FP    FN       25
#> 4 1          1      TN    TP       43
```

The basic calculation then will look in the following way:

$$\frac{TP + TN}{FP + FN + TP + TN} = 1 - Error\ .$$

True positives are data points labeled as positive that are actually positive whereas false positives are data points labeled as positive that are actually negative. True negatives are data points labeled as negative that are actually negative whereas false negatives are data points labeled as negative that are actually positive.

$$\frac{TP}{TP + FP}\ \left(\frac{TN}{TN + FN}\right).$$

As a standard for classification method we will use metrics such as F1-score, precision, and recall. Instead of giving preponderance to either recall or precision, we use F1-score to essentially combine them into one metric. Precision in our case examines specifics of one class. For instance it would mean that the model will become better at predicting the fraudulence of a given transaction/address rather than non-fraudulence. In general, Recall metric refers to the extraction of a correctly predicted result by the model. Therefore, the Recall metric can be deemed to be a kind of purveyor of data points of interest.

$$\frac{TP}{TP + FN}\ .$$

The use of different methods for classification will allow us to yield far more reliable results and train our model to achieve higher specificity and accuracy. One of the models that will help us to build an efficient classification model is Random Forest. The gist of Random Forest has arbitrarily selected *k* features from the total *m* features, in cases where $k < m$. The Random Forest randomly selects k features from m total features, where k < m. From the entirety of k features, it calculates the node point *d* by alluding to the best splitting spot. Afterward, the nodes are divided into smaller nodes, again by using the splitting spot. In general, this iterative process is executed until *l* number of nodes has been reached. The Random Forest is built by reiterating the same process *n* number of instances and creating *n* number of trees. The results are obtained by the modal value of categories obtained by individual trees.

The main hurdle in establishing a fully working machine learning model is to train it on as much as possible a diverse set of scenarios. In the case of fraudulent transactions and their not as prolific (fortunately) scope, a case can be made that a system or model created on the sparse set of scenarios may not glean a sufficiently adequate result. This can partially be remedied with the help of historically available data on fraudulence as well as pre-imputed datasets. We also resort to the binary classification model for the time being as it allows us to minimize the computational overhead and simplify the model for the very first iteration while in the database-building phase.

System Governance and Ability to Shape Cybersecurity Space Together

On the path of creating a decentralized autonomous system that will operate on the basis of community-defined development and subsist on the communal efforts, it’s mandatory to establish a type of governance protocol that will feature and incentivize interconnectedness. HAPI Protocol first and foremost is a neutral observer and reporter that is fitted according to the predetermined configurations of an entity using it. The protocol can be augmented in various ways that would define the all-encompassing institution of regulatory means (AML, CTF) and appendment of illicit addresses and unlawful liquidity manipulation into the database.

System governance will include adding new organizations of different kinds (incident reports, fund trackers, possibly government structures that can sanction individual's assets based on international laws) to network, management of reward structure, and case-by-case courts.

The governance structure of HAPI will consist of a committee. The committee is an inner circle of decision-making. HAC (HAPI Authority Committee) is one of the cogs in the governance body that consists of members of the HAPI team and external trusted third parties that are responsible for proposal execution.

The main voting power within the protocol is concentrated in the native token of the protocol - HAPI. HAPI token essentially is used as a vote delegation tool and each participant is able to delegate their vote accordingly.

Each proposal can be divided into stages and types. Each stage defines the process of approval. Every type reflects the complexity of an alteration in question.

Stages:
PROBE - low threshold voting to detect general interest in a proposal.

IMPL - medium threshold voting for a particular set of changes that should be implemented.
DEPLOY- high threshold voting to deploy changes to the smart contract.

Type 2: Adjusting smart contract configuration
Proposals of type are responsible for basic smart contract configuration adjustments: reward rates, thresholds, etc.

Stages:
PROBE - low threshold voting to detect general interest in a proposal.
DEPLOY - high threshold voting to apply changes to the smart contract configuration.
Type 3: Reporters and data providers management
This type of proposal is very important for the ecosystem to welcome new data providers and punish the bad agents.

Stages:
DEPLOY - the decision is made in a single stage with a previously decided threshold.
Type 4: Data regulations
This type of proposal does not impact on-chain infrastructure directly but governs Oracle data sourcing algorithms and approaches, as well as the support of different networks.

After a successful vote on a proposal, it is up to the HAC to enroll the proposed changes to the production network within a two-week period after the final proposal vote.

Each participant is able to submit data on the address deemed to be fraudulent. By issuing a proposal and supplying it with enough evidence of a given address being related to the malicious activity, a consummated verdict can be enacted. If a verdict is positive and the proposal is assented to, the address is moved to the database of HAPI SC and flagged as risky. Submission of an address for examination requires a certain amount of HAPI to be proposed. This is implemented for two reasons. Firstly, to avoid a deluge of addresses and time-consuming effort of sifting through each and giving it a respective appraisal and, secondly, to ensure a reward structure for validators within the governance system.

Bibliography

1. *“12 Month Review Revised FATF Standards Virtual Assets.” Second-12-Month-Review-Revised-FATF-Standards-Virtual-Assets-VASPS, www.fatf-gafi.org/media/fatf/documents/recommendations/Second-12-Month-Review-Revised-FATF-Standards-Virtual-Assets-VASPS.pdf.*
2. *“Are Centre Centralized Cryptocurrency Regulations the Answer? Are Cryptocurrecny Regulations the Answer? Three Countries; Three Different Directions .” Brooklynworks, brooklynworks.brooklaw.edu/cgi/viewcontent.cgi?article=1960&context=bjil.*
3. *Carrillo, Casey. “‘Probably the Largest Kyc Data Leak in History’ Demonstrates the Importance of Bitcoin Privacy.” Bitcoin Magazine: Bitcoin News, Articles, Charts, and Guides, Bitcoin Magazine: Bitcoin News, Articles, Charts, and Guides, 29 Mar. 2021, bitcoinmagazine.com/business/probably-the-largest-kyc-data-leak-in-history-demonstrates-the-importance-of-bitcoin-privacy.*
4. *“The Chainalysis 2021 Crypto Crime Report.” Chainalysis, go.chainalysis.com/2021-Crypto-Crime-Report.html.*
5. *“Crypto Compliance.” Crystal Blockchain Analytics for Crypto Compliance, crystalblockchain.com/investigations/how-crystal-investigations-truly-make-a-difference/.*
6. *“Crypto Crime.” Crypto Head, 5 July 2021, cryptohead.io/research/crypto-crime/.*
7. *“Crystal Blockchain Analytics 2021.” Crystal Blockchain Analytics for Crypto Compliance, crystalblockchain.com/investigations/darknet-interactions-bitcoin-a-crypto-activity-analysis-for-may-2021/.*
8. *Eisenbeis, Robert. “TNB and the Regulatory Dialectic.” Accelerating Progress, 14 Dec. 2018, www.spglobal.com/marketintelligence/en/news-insights/trending/bh8zuqyjenucv2z_azcaza2.*
9. *“EU's AMLD5 Is Not Enough.” Bruc Bond, 3 Jan. 2020, www.brucbond.com/article/eus-amld5-is-not-enough.*
10. *Hinteregger, Abraham, and Bernhard Haslhofer. “Short Paper: An Empirical Analysis of Monero CROSS-CHAIN TRACEABILITY.” Financial Cryptography and Data Security, 2019, pp. 150–157., doi:10.1007/978-3-030-32101-7_10.*

11. *Huang, Danny Yuxing, et al. “Tracking Ransomware End-to-End.” 2018 IEEE Symposium on Security and Privacy (SP), 2018, doi:10.1109/sp.2018.00047.*
12. *Huang, Sherena. “Cryptocurrency and Crime.” FinTech, Artificial Intelligence and the Law, 2021, pp. 125–143., doi:10.4324/9781003020998-11.*
13. *Perrin, Chad. “The CIA Triad.” TechRepublic, TechRepublic, 30 June 2008, www.techrepublic.com/blog/it-security/the-cia-triad/.*
14. *Tracing Transactions across Cryptocurrency Ledgers - Usenix. www.usenix.org/system/files/sec19-yousaf_0.pdf.*
15. *E. Ben-Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza. Zerocash: Decentralized anonymous payments from bitcoin. In 2014 IEEE Symposium on Security and Privacy, pages 459–474, Berkeley, CA, USA, May 18–21, 2014. IEEE Computer Society Press | S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage. A fistful of bitcoins: characterizing payments among men with no names. In Proceedings of the 2013 Internet Measurement Conference, pages 127–140. ACM, 2013*