Decentralised Gambling: The York Combined Transaction Set

Oliver James Scholten
Department of
Computer Science
University of York, UK
ojs524@york.ac.uk

David Zendle
Department of
Computer Science
York St. John University, UK
d.zendle@yorks.j.ac.uk

James Alfred Walker
Department of
Computer Science
University of York, UK
james.walker@york.ac.uk

ABSTRACT

This paper describes the York Combined Transaction Set (YCTS), which offers a single consolidated list of publicly available gambling related transactions derived from the Ethereum blockchain. This data includes over 1.4M individual transactions across 17,000+ unique addresses, which represent spending on decentralised gambling smart contracts. These contracts, and corresponding applications, have been selected based on their popularity as presented by an officially recognised ranking service, and have transacted over £240M by naïve estimation over the past 10 months. Given the historical opacity of data driven gambling research, our contribution is to identify, and make available in a simple form, the transaction data found on the Ethereum blockchain such that existing questions surrounding player spending can be explored, and differences between decentralised and traditional forms of gambling can be identified.

I. INTRODUCTION

Data driven gambling research has historically been limited by a lack of data availability, and limitations on its use in published work. Decentralised gambling, or gambling using cryptocurrency, presents a paradigm shift in these limitations, offering publicly accessible transaction level data on gambling activity across growing groups of individual user addresses. Such platforms have historically not been large enough to warrant mainstream academic attention, especially in comparison to more traditional gambling providers (e.g. internet gambling and physical casinos). However, this problem of scale, and therefore analytical generalisability, has recently been challenged by the creation and rise of several decentralised gambling applications built upon Ethereum smart contracts. This paper aims to briefly describe such contracts, and presents a large scale ($n=17,046$) consoli-

dated data set derived from the publicly available transactions which use these contracts.

Decentralised Applications

Decentralised applications operate using cryptocurrency networks by interacting with ‘smart contracts’ [1]. These contracts are computer programs which can be executed by transferring funds to the cryptocurrency address with which they are associated. Conceptually this is equivalent to purchasing a service in the real world - funds are initially transferred, and a previously agreed action is performed. Smart contracts replace these typically human-based agreements with software, removing dependence on third parties. This independence invites the creation of a wide variety of decentralised applications, from censorship resistant social networks, to video games, to gambling services. In the case of gambling services this mode of operation poses a significant problem for regulatory bodies and law enforcement given the pseudo-anonymity of users, and their international availability outside of state limitations on the use of cryptocurrencies.

The risks that decentralised applications pose to existing gambling platforms, consumer protection measures, and regulatory ability, have only recently been recognised [2]. These risks stem from the *unstoppable* nature of the smart contracts in question (given their global distribution), and the pseudo-anonymity of those initiating transactions on the network.

Finally, links have been identified between cryptocurrency trading and risk of problem gambling, depression, and anxiety [3]. Although this work does not explicitly address the use of decentralised gambling services, it touches on several related issues, specifically the rise in use of cryptocurrencies generally, and their relation to risk taking behaviour. As all users of decentralised gambling services must have, by definition, traded cryptocurrency, or at least purchased it for real world money through an exchange, findings from this area may equally apply in this even younger domain. With this considered, we briefly describe the layout of this work, then go on to describe smart contracts - the underlying technology in decentralised gambling - in greater detail.

The layout of this paper is as follows; Section II introduces smart contracts through the lens of gambling provision. Section III describes the process of gathering data from these contracts, including the raw data gathered and cleaning steps applied before consolidation. Section IV provides a description of the cleaned data, including interpretations of the data with respect to gambling activity. Section V describes how the data set can be accessed, including links to the corresponding Open Science Framework project. Section VI then discusses the limitations of the data described in this work, and briefly describes how they may be overcome in future work. Section VII then discusses wider issues with respect to the transaction data available and limitations on potential findings. Finally sections VIII and IX provide a final summary and future avenues of research using this new data set.

II. SMART CONTRACTS

The software contracts discussed in Section I are commonly known as smart contracts [1]. A smart contract is created by writing a software script and transacting with it on the Ethereum blockchain. Once created, these contracts will exist indefinitely and can be transacted with by any other address on the network, making them instantly global, and unstoppable without destruction of the entire network on which they exist. Decentralised gambling providers, as briefly described in Section I, can be described as contracts which accept payment and perform chance based calculation, returning or not returning payment depending on the outcome of that calculation. This description draws parallels with that of slot machines [4] and similar gambling devices, whilst meeting all of Griffiths’ criteria for distinguishing gambling from simple risk taking [5].

Smart contracts, by nature of their design, are censorship resistant; in that no individual entity may prevent others from transacting funds to them - thereby executing their code. Further, in the context of gambling services, protective measures put in place for those exhibiting problematic gambling behaviour cannot be applied, as the identity of the player is unknown, and the contracts they are interacting with cannot be altered.

III. DATA PROCESSING

This section describes the nature of the data collected and the subsequent transformation and cleaning steps performed to extract the meaningful transactions from the underlying smart contracts. In the context of decentralised gambling transactions, we define meaningful as ‘initiated by a human, likely to be a bet’. The implementation of this definition is discussed following a description of the raw data.

Raw Data

The raw data gathered as part of this work includes the full transaction histories of each of the addresses listed

York Gambling Set				
From	Cost	Value	Provider	Time
=====	=====	=====	=====	=====
=====	=====	=====	=====	=====
=====	=====	=====	=====	=====
...				

Figure 1. Final structure of cleaned and consolidated transaction data.

in Table 1. Each of these addresses is publicly affiliated with a website, which acts as an *optional*¹ interface through which transactions can be initiated. These interfaces are cosmetically equivalent to games found on internet gambling sites, but only act as ‘skins’ for the single transaction which underpins the betting. The addresses with which these sites are affiliated can be found by following transactions on their leader boards *if available*, or are openly displayed on the pages themselves. The complete raw data spans 2.7M transactions, 1.6M of which come from the leading smart contract (provider 1 in Table 1).

Cleaning

For each of the transactions in each of the three data sets, all transactions where the ‘eth’ column exactly equalled zero were removed. This follows the definition of ‘meaningful transactions’ above, given the improbability of a human initiating a zero value transaction. This removed 1,303,887 transactions from the existing 2.7M, or a 48% reduction across all data. This reduction is expected given the utility of zero-value transactions in the smart contracts selected. These *non-meaningful* transactions are generated by the smart contracts themselves and range in function from verifying random number generation to calculating winnings and other ‘utility’ functions.

With non-zero transactions now isolated, the addresses of each of the three sets can be preserved by the addition of a new ‘provider’ column, which contains a unique integer (between sets) such that when sets are combined their origin is identifiable. This step added an integer for each row (1, 2, or 3), depending on the provider as found in Table 1. This integer based method was chosen over simply repeating the provider’s address multiple times given the reduced storage size of a single integer versus a 20 byte string, especially given the number of transactions.

The final step in the cleaning process is to merge the three sets described above, and rename/reorder columns from those presented by the Etherscan API to more intuitive values. This results in a single table of 1,420,292 rows, with columns {from, cost, value, provider, time}. We now describe each of these columns, and their interpretation in the context of decentralised gambling.

¹These interfaces can be considered optional as the related smart contracts can also be executed by manually transferring funds from a cryptocurrency wallet.

Table 1. Addresses used in composition of YGTS. Dates in DD-MM-YYYY format. Gathered 22/05/2019.

Provider	Address	Date Created	Transaction Count
1	0xD1CEeeee83F8bCF3BEDad437202b6154E9F5405	07-09-2018	1,651,735
2	0x999999C60566e0a78DF17F71886333E1dACE0BAE	10-12-2018	557,812
3	0xA52e014B3f5Cc48287c2D483A3E026C32cc76E6d	04-08-2018	514,632
Total:			2,724,179

Table 2. Total transaction values, in ETH, for each provider. Unique addresses are counted first using a naive summation (striked out) and second across providers yielding a lower value as duplicates are removed.

Provider	Non-zero Transactions	Unique Addresses	Total Value
1	886,836	4,574	557,120.97
2	278,993	11,273	861,207.75
3	254,463	2,200	302,174.99
Total:	1,420,292	18,047 17,046	1,720,503.72

IV. DATA DESCRIPTION

Figure 1 shows the structure of the consolidated data set described in Section III. The first column contains the cryptocurrency address from which the transaction was initiated. This may be interpreted as a unique individual identifier under the assumption of a single address per user. This assumption relies on the lack of benefit in creating multiple addresses per person for the purposes of sending funds to the smart contracts selected, a real world comparison being an individual registering two bank accounts to purchase the same thing. The ‘cost’ column contains an integer representing the amount spent on verifying the transaction itself on the cryptocurrency network. For the purposes of gambling spending analysis, this may not be as useful as following columns, as the network transaction cost may not relate to bet size, but is included for completeness of user spending. The third, and most important column, contains the value of the transaction made to the smart contract, which is represented as an integer in the fourth column. This can directly be interpreted as the amount of cryptocurrency wagered on a single gambling activity, which may manifest as a dice roll or coin flip in the interfaces described in Section I.

All of the values in the YCTS remain in terms of the underlying cryptocurrency the transactions used, in this case Ether - the name of the cryptocurrency used in the Ethereum network. Taking an approximated estimate of the value of one Ether to be £140,², the naïve total value of gambler spending in the York Combined Transaction Set comes to £240M.

V. DATA ACCESS

The data described in this work is available for download at <https://osf.io/dhcyq/>. The data set in its entirety requires approximately 118MB of storage, and is stored in CSV format. The linked project also contains

an itemised version of the description in Section IV for convenience.

The data presented is also a subset of the already publicly available Ethereum blockchain, gathered using the [Etherscan API](https://etherscan.io/)³. This data can also be accessed by directly interacting with an up-to-date Ethereum node, although an API is recommended for simplicity.

VI. LIMITATIONS

The data described in this work contains only the bets made by anonymous users on the Ethereum cryptocurrency network to known decentralised gambling contracts. This means, that although transaction histories are available for each individual address, they are not included here, meaning questions surrounding which portion of a user’s wallet are being spent on gambling, and whether or not individual bets were a win or a loss, cannot be answered without further data gathering. Further gathering of this type is however possible, given the public availability of all transactions on the network, and uses identical data access methods to those discussed herein.

Second, although the data set contains over 17,000 unique addresses, there is no way to verify that each of these addresses is being operated by a single person. In this vein, we assume each address to represent an individual, however no further data gathering can confirm or deny this assumption. This limitation must also be considered when augmenting this data set with others as described above. This is especially important should conclusions with significant real world repercussions be drawn, as discussed in the following section.

Finally, given the ongoing and unstoppable operation of the smart contracts presented here, the amount of data made available will only increase over time - both with respect to the three contracts identified in this work, and

²<https://coinmarketcap.com>, accessed 23/05/2019.

³API available at <https://etherscan.io>, last accessed 11/07/2019

others yet to be discovered. This means that although this work makes available a large data set for academic use, the tools used to gather such data represent a more powerful contribution, so will be made available following further work on their development.

VII. DISCUSSION

Given the rich, transaction level gambling spending data available, a wide range of analytical approaches and high level questions appear. These include the application of time series analytical methods capable of identifying temporal trends in spending, to network topological investigations as found in complex network research. Such analysis may answer questions, such as how typical user spending varies over time, whether ‘sprees’ of gambling activity exist, and how users of the same decentralised gambling platform interact financially, if at all.

Similar recent work includes investigations on disproportionate spending and problem gambling [6], whereby problem gamblers may be identified within this data set using analysis of their transaction frequency and magnitude. If possible, this represents an immensely useful area of investigation as the automatic identification of those at risk of developing problematic gambling behaviours may allow preventative steps to be taken - as opposed to treatment of already established behaviours.

VIII. CONCLUSION

This paper has briefly discussed, at the technical level, the nature of several existing decentralised gambling smart contracts, and presented consolidated data relating to their use. Interfaces to these contracts, whilst themselves only executing code when transacted with, include designs similar to gambling services as provided by internet gambling companies. The data gathered is a subset of the publicly available Ethereum blockchain, where transactions relate to addresses with public affiliation with decentralised gambling providers. The cleaned and consolidated version of this data has been made available for download, bypassing the programmatic tools required for their extraction.

IX. FURTHER WORK

The data set presented in this work includes the decentralised gambling activity of over 17,000 accounts at the transaction level. Future exploration of this data set may generally fall under two main objectives. The first being the validation of existing theories on gambler spending and similar topics under the assumption that the similarity of interfaces makes such analyses comparable. The second being the likewise comparison of data sets gathered in other, more traditional gambling services, such that any changes in gambler spending between services can be identified. Each of the goals above hold a huge number of variations in the scope of the underlying work, and in the analytical approaches required for their completion.

Further, the data presented here, whilst powerful when analysed in isolation, will likely benefit greatly from complementary data sets. Examples of such sets may be the transaction histories of each account, transaction histories of internet gambling sites for comparison, etc. All future work in this area presents the possibility of extremely valuable findings, especially when framed in terms of identifying, assisting, and treating those exhibiting problematic gambling behaviour.

Disclosure

The authors declare no actual or potential conflict of interest in relation to the work presented or related data.

Acknowledgements

This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games & Games Intelligence (IGGI) [EP/L015846/1] and the Digital Creativity Labs (digitalcreativity.ac.uk), jointly funded by EPSRC/AHRC/Innovate UK under grant no. EP/M023265/1.

REFERENCES

- [1] V. Buterin et al., “A next-generation smart contract and decentralized application platform,” white paper, vol. 3, p. 37, 2014.
- [2] S. M. Gainsbury and A. Blaszczynski, “How blockchain and cryptocurrency technology could revolutionize online gambling,” Gaming Law Review, vol. 21, no. 7, pp. 482–492, 2017.
- [3] D. J. Mills and L. Nower, “Preliminary findings on cryptocurrency trading among regular gamblers: A new risk for problem gambling?,” Addictive behaviors, vol. 92, pp. 136–140, 2019.
- [4] N. D. Schüll, Addiction by design: Machine gambling in Las Vegas. Princeton University Press, 2014.
- [5] M. D. Griffiths, “Is the buying of loot boxes in video games a form of gambling or gaming?,” Gaming Law Review, vol. 22, no. 1, pp. 52–54, 2018.
- [6] I. Fiedler, S. Kairouz, J.-M. Costes, and K. S. Weimller, “Gambling spending and its concentration on problem gamblers,” Journal of Business Research, vol. 98, pp. 82 – 91, 2019.