THE JBBA

PEER REVIEWED RESEARCH

OPEN ACCESS
ISSN Online: 2516-3957
ISSN Print: 2516-3949
https://doi.org/10.31585/jbba-2-2-(6)2019

Blockchain Investigations - Beyond the 'Money'

Simon F. Dyson
NHS Digital, Leeds, U.K
Correspondence: simon.dyson@protonmail.com

Received: 18 July 2019 **Accepted**: 7 August 2019 **Published**: 13 August 2019

Abstract

Cryptocurrency investigations have centered almost entirely around the transfer of value "money" or a cryptocurrency asset. The use of cryptocurrency for illicit purposes, especially Bitcoin, is well documented both in academic writing, media reporting and even film documentaries. The infamous SilkRoad marketplace in addition to the millions of dollars spent within dark markets on drugs, guns and assassinations have grabbed the headlines. This paper looks at how blockchain is creating new areas of investigation that are yet to be explored in detail. This scenario-based research examines the hosting of stolen data (P.I.I) personal identifiable information on a distributed blockchain host where the data is also accessible. The platform used is based on Ethereum infrastructure but demonstrates just one available platform that poses the paradigm. The paper examines the considerations through the lens of an incident responder /cyber investigator, forensics examiner and data controller. The scenario highlights distinct differences in considerations from a traditional response compared to dealing with the immutable and unstoppable distributed technology. The paper concludes that more is needed to be done to understand digital forensics in the blockchain era and the need to develop beyond track and trace in the cryptocurrency investigative toolbox. The discussion also brings forth how data retention and GDPR requires consideration when applying it to blockchain systems.

Keywords: *Blockchain, Distributed-hosting, Distributed-storage, Ethereum, Swarm, Forensics*

1. Introduction

Research into cryptocurrency has focused generally on the transfer of value. The use of cryptocurrency in large scale criminal activities is well documented in cases such as the Silk Road drugs marketplace or in large ransomware campaigns such as Wanacry. The focus has been on the "follow the money" aspect in order to locate the perpetrators. The underlying technologies have however developed since the inception of Bitcoin in 2008. Blockchain technology is now scaling and developing new features now able to support multiple data and communication protocols across its stack. Law enforcements focus has remained around the large cryptocurrencies however the use of smart contract technology and now distributed computing and storage creates a new set of problems for investigators and those responding to incidents. This paper sets out a common leak of personally identifiable information (P.I.I) where it is hosted on blockchain technology and how the traditional responses are required to adapt. The scenario uses Ethereum and its related technology to host files. There are a number of cryptocurrency/blockchain assets that can host the data in a similar nature. A distributed blockchain by its design contains properties that are not inherent in traditional hosting services. A blockchain is immutable in general terms so they are unstoppable and have no central authority or body.

2. Scenario and Roles

The scenario is to replicate the discovery of files taken containing (P.I.I) personally identifiable information from a server and hosted externally. The hosting, however, will take place on a distributed blockchain system. In order to establish if the PII information is legitimate, a comparison will need to take place, this will entail a visual comparison of the data. A forensic comparison of the data will need to be conducted using traditional methods to hash the file contents and examine EXIF data contained within the file. Cyber investigators searching for online hosted material will examine records of web hosting companies to see the I.P data for the hosting company and registrar details such as WHOIS information. Data controllers hold the responsibility for the holding storage and protection of the data. The data controller will need to make decisions about steps that are possible to minimise the damage. Each role will respond using traditional methods and record the findings. A discussion section will reflect on the approaches and highlight quick wins and areas that require further work.

2.1. Cyber Investigator / Incident Response

This role will respond to initial reports and record and utilise OSINT Open Source Intelligence sources to discover evidential information to assist the investigation. The coordination of tasks to systems administration for internal log investigation and other closed source materials will be conducted.

2.2. Digital Forensics

Examination of digital material will be conducted by the digital forensics team member. This will include host forensics and also comparison of highlighted online material where required. They will take a forensic analytical approach in order to approach the problem.

2.3. Data Controller

As the responsible owner of the data, the controller will be consulted on the state of the investigation. The controller will establish additional tasks that would assist to protect the data or prevent further dissemination.

3. (GDPR) General Data Protection Regulation

In May 2018, the General Data Protection Regulations came into effect and incorporated existing legislation to protect people's data and their rights. GDPR is covered in depth in numerous resources so in this section a focus on some key themes that will be later visited will be briefly documented. GDPR covers data that belongs to people who are in the GDPR aligned nations, Europe and some additional territories. The rules outlined cover those entities that are considered a data controller or processor. A controller is the entity that holds the data for purpose, and they will process for their agreed business requirements. A processor is considered a third party that is doing something with the data on behalf of the controller, an agreement will define what that process is. GDPR defines that personal data can generally identify somebody or be used for that purpose and it offers protection to that data. There are also additional protections to sensitive personal data that protects special characteristics. The term PII (Personal Identifiable Information) is not defined by GDPR but is commonly used and will be covered as personal data under GDPR and this scenario. There are 8 individual rights that are listed under the new act. These rights are; the right to be informed, the right of access, the right of rectification, the right to erasure, the right to restrict processing, the right to data portability, the right to object and rights of automated decision making and profiling. The right to erasure is one of the more complex and powerful rights that is created in the new act. This right caused many to question the ability of blockchain to function under such a regime. The conflict of immutability as an absolute property of blockchain in comparison to the legal requirement to deletion of GDPR is cited as pushing solutions to standard databases[1]. There are other potential ways forward such as gaining consent for perpetual processing. It is argued that address hashing is pseudonymous and that the effort to de-obfuscate a hash is disproportionate so would stand as it would not likely identify an individual. Permissioned blockchains are also suggested in order to control the data but they don't fit the public and permissionless systems of large cryptocurrency structures. There are also systems using new encryption methods such as zkSNARKS and RingCT methods that could protect data throughout the complete process [2]. Tokenised solutions are appealing although they may require off-chain processing but the use of distributed storage is possible through Ethereum – SWARM or IPFS[3]. The use of a smart contract with an upgradable contract section could allow amendable content but record the transaction metadata and deletion process[4]. Implemented correctly the ability to control and make accountable sharing structures with blockchain could strengthen systems to comply with GDPR.

4. Decentralised Blockchain Storage

Decentralised networks have been utilised for numerous cryptocurrency projects with the ability to trade tokenised value they have become used for a new wave of "Digital money". Blockchain technology itself has evolved behind the headlines of boom and bust price fluctuations and Silk Road drug dealing dark markets. The introduction of Smart Contracts utilised by Ethereum and now other blockchain technologies allows Turin complete languages and sections of code to produce complex computational outputs. Using resources on Ethereum for example is expensive if you process through the Ethereum Virtual Machine (EVM) the world computer, each byte and code execution has a price to pay using "gas". Utilising "gas" small amounts of the currency this ensures that the "halting problem" is addressed and a denial of service attack or forever loop will be too expensive to conduct. There are however a number of blockchain projects that are looking to use an additional protocol or system to provide blockchain storage using peer to peer nodes incentivised to the system. The creation of a decentralised storage system solves a number of computing problems, it creates resilience as files are striped across multiple nodes in a system. The ability to reside on multiple nodes reduces single points of failure or risk from physical events such as earthquake, tsunami or power outages. A decentralised system uses nodes in the control of world users who are incentivised to "mine" or provide a service similar to miners and Bitcoin nodes. Services such as Dropbox operate a storage system that allows a cloud storage system however the service is a centralised under one organisation. The company is subject to US law, so privacy therefore is not guaranteed as the ability to access, subpoena, court order and secret service oversight. Nodes in a decentralised generally hold only partial fragments of the file so physical integrity is maintained as the file portion is fragmented and optionally encrypted. There are a number of decentralised file storage systems namely, IPFS (Inter Planetary File System) developed by Protocol labs this part of the system allows for distributed storage, Filecoin [5] is an additional service to incentivise storage by paying miners to store. IPFS as a protocol is used by a number of other projects and is cross blockchain agnostic [6]–[8]. In addition to the above there are other distributed storage projects in various phases of production these include Storj, Sia and Maidsafe [9]–[11]. Ethereum has its own sub project called "Swarm" this will be explored in the next section.

5. Ethereum Swarm

Swarm was designed to create a system to store dapp (Decentralised Application) code, resources on a peer2peer system. The ability to access material outside of the Ethereum chain reduces the cost of storing larger files or code in a smart contract off chain where it is cheaper to store. Dapps by their nature are applications that are not a singular stored item, therefore the use of larger code sets and files to produce more complex and visually focused items require more storage. The ability to access resource from the Swarm protocol layer allows this exchange maintaining a fully decentralised eco-system. The system will maintain the properties of a truly decentralised system transaction layer on chain and storage another chain. This makes it non-censorable, fully redundant / resilient, DDOS resistant, highly available and secured by encrypted cryptographic signatures. Ethereum integration is used with a Swarm node and a Geth node, Geth is a "GO" programming language implementation version of Ethereum. This scenario will utilise both Ethereum Geth and Swarm working on the Ropsten test net, the closest to the production service. Ropsten allows integration with the services as if it was connected to the Mainnet where the technology is already live, with the advantage of not costing real Ethereum and "gas" to test and operate. Geth version (1.8.20-stable) and Swarm version (0.3.-stable) [12], [13].

6. Blockchain Domain Naming

The (DNS) Domain Name System is used to assist with searching the internet, it translates a human readable (URL) Uniform Resource Locator into the relevant Internet Protocol Address (I.P). This directs a query such as www.a_web_address.com to the root servers to the (TLD) Top Level Domain and to the domains name server that holds the record of the I.P address example 8.8.8.8. The ability to store domain naming information on a blockchain has existed for some time with services such as Namecoin offering various services including a name resolution stored on blockchain. The criminal use of decentralised DNS services does exist but is not extensively used [14], [15], [16]. The discovery of a recent botnet that was discovered to be cleaning up bad botnets was observed in the wild using Emercoin's distributed DNS implementation [17].

The (ENS) Ethereum Naming Service provides similar functions to a DNS system and is held and operated over the Ethereum blockchain [18].

7. Method

In order to replicate an intrusion event a number of files with identifiable meta-data will be created and hosted on a virtual machine. A base forensic examination will be completed to display (Modified, Access, Created) MAC date and times, Meta data that may also include geo-data serial number or other EXIF data. The scenario host is a machine running Windows 7, the host contains a folder on the desktop entitled "Work_items" containing related documents. The documents include an image of a passport that is used for identification of the customer in this scenario and contains PII information. In addition to the image are further documents including an XLS and CSV files, this contains customer details including PII data.

Scenario – a message is received by phone that a leak of company information has occurred, and a website URL is provided.

7.1. Investigation phase

7.1.1. Cyber investigator / Incident response

The cyber investigator is initially passed information provided by a telephone call that states the web URL hosting the company's potentially stolen information. The URL is placed into a browser on a standalone environment to ensure the reported event is not a social engineering ruse and to protect the main corporate network from malicious activity. Initial activity will ensure the link is live and that the data appears to be present, accurate recording of event will take place including a screen capture of the page. A capture of the page and the source code alongside an abstraction of pertinent files will be completed for further analysis. Data will be needed to be compared to corporate data to ensure the attack is a legitimate attack and is not a hoax. OSINT Open source intelligence will reveal additional information about the web hosted material. The source code may reveal hosting details or frameworks used to create the site, these may include author and other meta data of interest to the cyber investigator. Source code can also reveal other links hosted on other sites or resources that may allow additional investigative leads. As discussed in Open source intelligence techniques by Michael Bazzell there are numerous services that assist in the location of a website these include some of the following important areas [19]:

- Protocol
- Website name and top-level domain information
- I.P address
- Whois
- Registration data
- e-mail addressing
- The hosting company (Server hosting)
- Domain hosting (Name holder)
- DNS zone tranfers
- Registrar change history
- Ad-sense / analytical tokens – numbers
- Robots.txt
- Shodan

7.1.2. Digital Forensics

Following information from the original call the forensic response team will react to the main areas of data storage. The firewalls and server logs will be checked for intrusion or indicators of compromise. The data storage servers will be examined, and a RAM dump will be executed on each device. This will capture processes, network connections and master file table entries that will enable initial triage to identify any breach information. Identification of the information can take place by using methods such as hashing values and searches for names or data from the leaked source to discover if the information is owned by the company.

The order of volatility is Processor, Network, Main Memory, Semi-volatile, Resident data, Remotely logged and any data on archival media [20]

In this scenario, live data should be considered before a raw dump, if the memory dump crashes then the machines critical live data could be lost. A memory dump should be obtained and analysed the machine can be shut down and retained for a full forensic image if required.

The forensic response to an incident would record the process using contemporaneous notes and photographs.

Examine with the visual inspection of a machine and examination of live desktop activity

- Live data – command line – time & date, network connections (netstat), current user, tasklist
- Memory RAM capture – full, Dumpit.exe
- Any operational/incident specific investigation tasks.
- Power down machine when examination complete retain for full disk imaging.

7.2. Operational Phase

7.2.1. Data Controller

The General Data Protection Regulation GDPR introduced in 2018 enforces businesses and those who control data to protect the rights of the citizens whose data is held. Each European country or participating country must introduce a body to monitor and administrate enforcement of fines and breaches of the code. In the U.K the ICO (Information Commissioners Office) hold this position, they provide guidance, advice and are the primary contact if a breach occurs. GDPR requires a company that is aware of a breach of personal data to report to the supervisory authority in the U.K to the I.C.O within 72 hours of been aware that a breach has occurred. Where it is likely that a breach will affect the rights and freedoms of the individuals on who the data relates then they must also be informed "without undue delay" [21]. The principles that are to be considered around data are the security triad of Confidentiality, Integrity and Availability. The rights of the data subject are to be considered and notification made if the breach is likely an adverse effect on the data subject. An example would be where full personal data and financial data are lost these are likely to incur subsequent fraud offences using the identities of the data subject [22].

GDPR therefore requires all companies that process data in the EU or about people in the EU to have policies and procedures to detect, investigate and report on incidents with accuracy.

GDPR has a number of requirements in relation to information to be provided to the supervisory body in response to a breach. The below section details the requirements and these points will be addressed in the breach investigation plan for the data controller.

- Description of the personal data (data categories, number of individuals, number of records)
- An assessment on potential consequences following the breach
- Following the breach what measures have been or will be taken in order to mitigate risk and harm following the breach. (ICO GDPR breach guidance [23]).

It is obvious from the above requirements that a response from the cyber investigators / digital forensic team is essential in providing timely and accurate reporting to ensure the data controller can make informed decisions on the subject.

8. Initial Response

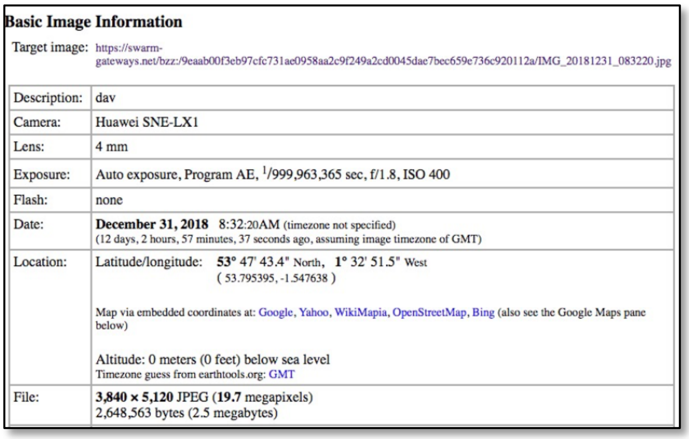
The scenario starts with a report of information reported into the Cyber security team. This was initially reported as a URL and the action will start by the teams who will perform incident response according to their response plans. The URL was reported as:
https://swarm-gateways.net/bzz:/9eaab00f3eb97cfc731ae0958aa2c9f249a2cd0045dae7bec659e736c920112a/

9. Findings

9.1. Cyber Findings

Initial actions resulted in the preservation of the online material and the capture of HTML information of the files hosted on the site. The site contained three items of interest an image of a passport and two data files for download. The nature of the message on the site suggested an insider threat this requires further internal work to attribute.

The image was reverse searched to see if the image was hosted elsewhere on the web, this was to establish if this was disseminated elsewhere or if it was a hoax using another source. The EXIF data was examined from the passport photo, this provided metadata that included time dates, make, model, image composition and crucially geo location, see Fig 1. This established the passport image was taken in a popular café used by the sales team to on-board new customers close to company premises. This provided metadata that allowed attribution in this circumstance in the scenario set out.



Basic Image Information

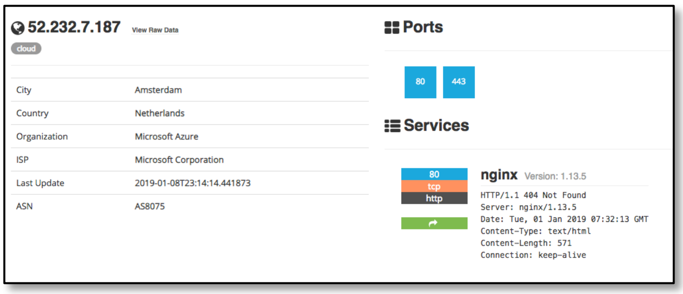
Target image: https://swarm-gateways.net/bzz:/9eaab00f3eb97cfc731ae0958aa2c9f249a2cd0045dae7bec659e736c920112a/IMG_20181231_083220.jpg

Description:	dav
Camera:	Huawei SNE-LX1
Lens:	4 mm
Exposure:	Auto exposure, Program AE, ^{1}/999,963,365 sec, f/1.8, ISO 400
Flash:	none
Date:	**December 31, 2018** 8:32:20AM (timezone not specified) (12 days, 2 hours, 57 minutes, 37 seconds ago, assuming image timezone of GMT)
Location:	Latitude/longitude: **53° 47' 43.4"** North, **1° 32' 51.5"** West (53.795395, -1.547638) Map via embedded coordinates at: Google, Yahoo, WikiMapia, OpenStreetMap, Bing (also see the Google Maps pane below) Altitude: 0 meters (0 feet) below sea level Timezone guess from earthtools.org: GMT
File:	**3,840 × 5,120** JPEG (**19.7** megapixels) 2,648,563 bytes (2.5 megabytes)

Figure 1. EXIF data from the passport image

9.2. OSINT – Findings

The domain was subject to a reverse look up, WHOIS search, and it revealed a hosted service. The information was shown to a Windows Azure instance based in the Netherlands. The domain registrar shows a named contact with addressing. The I.P address was established with versioning and port numbers on a Shodan scan that revealed web ports 443 and 80 were the only active services see Fig 2.



52.232.7.187 View Raw Data

cloud

City	Amsterdam
Country	Netherlands
Organization	Microsoft Azure
ISP	Microsoft Corporation
Last Update	2019-01-08T23:14:14.441873
ASN	AS8075

Ports

80 443

Services

80 tcp http

nginx Version: 1.13.5

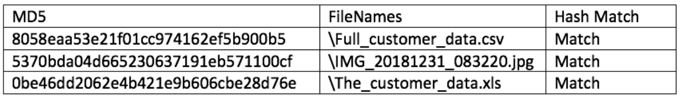
```
HTTP/1.1 404 Not Found
Server: nginx/1.13.5
Date: Tue, 01 Jan 2019 07:32:13 GMT
Content-Type: text/html
Content-Length: 571
Connection: keep-alive
```

Figure 2. Shodan results from the OSINT scan

9.3. Forensic Findings

The files recovered in the discovery phase were provided for forensic analysis. The items were hashed, and the metadata examined. This information enabled identification of the company database server where the data was likely to be stored. The forensic actions as previously described were enacted capturing live, ram and forensic level data. The company server was examined, and activity was discovered around the folder of interest using Volatility. Artefacts were found in the MFTPARSER and SHELLBAGS modules that allowed activity from MAC (Modified Accessed Created) times to create a timeline of suspect activity. In addition, access to the registry keys through the Volatility modules allowed the USBSTOR to show activity in the timeframe, giving make model and GUID for the suspect USB. Table 1 shows the hashing data and the matches.

Table 1. Shows the hash detail and if the hashing matched from the blockchain storage and the host system



MD5	FileNames	Hash Match
8058eaa53e21f01cc974162ef5b900b5	\Full_customer_data.csv	Match
5370bda04d665230637191eb571100cf	\IMG_20181231_083220.jpg	Match
0be46dd2062e4b421e9b606cbe28d76e	\The_customer_data.xls	Match

10. Data controller – next steps

In this section discussed is considerations of the data controller for the blockchain element and not the general actions of the controller, the data is personal and a referral to the ICO is required in the time frames set. The current actions would now look to reduce the spread of stolen data. Legal action against the hosting company or a complaint procedure to the hosting services would be sought. A powerful tool for removal of data is the Subject Access Request procedure where a data subject has rights under GDPR for enforcement. Legal proceedings are complex and civil claims can potentially disrupt or force servers to shutdown such services as detailed between PML vs unknown [24]. This shows the complexity and interactions that a hosting company can be pursued to reduce the impact and required to remove content under national and international law.

11. Initial conclusion standard response

The conclusion established from the above investigation at this stage are mixed. The captures have been performed to an adequate standard but there are some items that confuse the investigation. The domain and services discovered in the phase point to the "swarm gateway" a service allowing a pass through of web traffic to the Ethereum network. The Microsoft Azure server hosted in The Netherlands and the registrar name highlighted is a project lead on the Swarm service. The registrar and the WHOIS information all resolve to an unrelated subject not the true location of the data, just the portal to find it. It is important to relay that there is no information hosted on the server it is a HTTP/S proxy API that allows access to the Swarm network. There are other gateways such as https://ensgateway.com/ and IPFS specific gateways. What legal action can be taken against a portal that contains no data but provides access. Similar to that of a tor gateway or node allowing access to a darkmarket.

The forensic investigation however demonstrates that the files and data hosted on the Swarm system are not altered and retain important meta-data. The comparison shows that the integrity of the file is retained and the hashing value and EXIF data is retained when recovering from the Swarm network this confirms attribution for the company.

The investigation can conclude that the personal information has been taken from the company and this has been conducted by an individual with authority to access the service. The attack was conducted by exfiltrating data and removing it on a USB device this is a classic insider attack. The file is hosted on an unstoppable blockchain where no legal avenue exists to remove or request a cease and desist.

12. Blockchain investigations

The Swarm decentralised system operates using the URL scheme identifier as "BZZ:" the location of the file is designated by a Swarm hash or an ENS assigned domain such as "photoalbum.eth".

In the example, the ENS domain is assigned as "Unstoppable.eth" - Ropsten testnet and this resolves the content of the stolen items as examined previously to the swarm hash.

There are a number of components that work to resolve the addressing. At high-level an Ethereum registry that tracks the domains and sub-domains on the network. There are additional registrars that are involved in the hosting and reselling activities of ENS names. The Ethereum Naming Service is used to bid and retrieve a human readable address such as "Unstoppable.eth" and this is done using an Ethereum account. On successful allocation of the bid the name is under the control of the account and using Smart contract calls can be accessed and communicated with to set the requirements in the contract. The ENS record requires a resolver assigned that links the human readable name, name-hash, account or content to a resolver. There is a public resolver frequently used however custom resolvers are possible to create and likely to be adopted in some Dapps or other services. A name hash is used to represent the human readable name and is combination of cumulative hashing of domain and naming using Keccak hashing [18]. Fig 3 shows a walkthrough of the process.

In the example of the scenario the content hash was created by Swarm '9eaab00f3eb97cfc731ae0958aa2c9f249a2cd0045dae7bec659e736c920112a" this hash was used to search the Swarm node to retrieve the full content. The Swarm hash is created using a chunk hash function with a merkle tree, this is currently

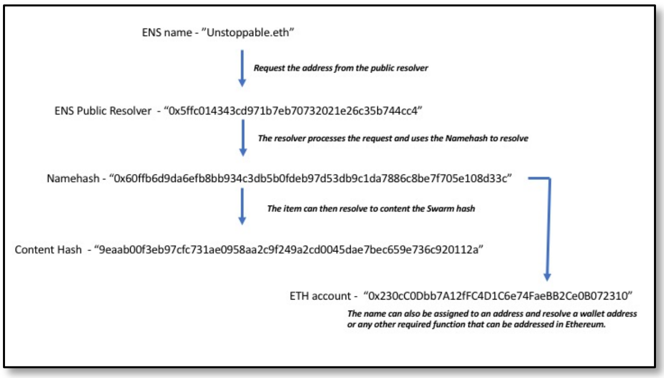
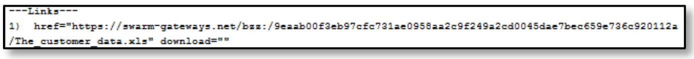

ENS name - "Unstoppable.eth"
Request the address from the public resolver
ENS Public Resolver - "0x5ffc014343cd971b7eb70732021e26c35b744cc4"
The resolver processes the request and uses the Namehash to resolve
Namehash - "0x60ffb6d9da6efb8bb934c3db5b0fdeb97d53db9c1da7886c8be7f705e108d33c"
The item can then resolve to content the Swarm hash
Content Hash - "9eaab00f3eb97cfc731ae0958aa2c9f249a2cd0045dae7bec659e736c920112a"
ETH account - "0x230cC0Dbb7A12fFC4D1C6e74FaeBB2Ce0B072310"
The name can also be assigned to an address and resolve a wallet address or any other required function that can be addressed in Ethereum.

Figure 3. A breakdown of the ENS and the different elements involved in name resolution

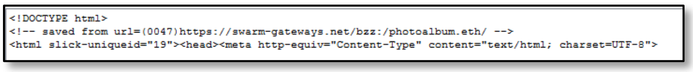
formulated using a 32 byte Keccak(256)SHA3. It is possible to create a hash of just a file or similarly in this case a folder with linked resources and files. In the meta-data for the html file the linked images are referenced as the hash and file Fig 4.



```
---Links---
1)  href="https://swarm-gateways.net/bzz:/9eaab00f3eb97cfc731ae0958aa2c9f249a2cd0045dae7bec659e736c920112a
/The_customer_data.xls" download=""
```

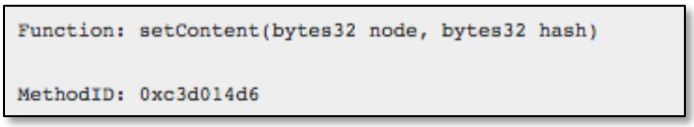
Figure 4. Web link that shows the Swarm hash in the HTML link data from the page

In Fig 5 displayed is a resolved address through the ENS service linking to the swarm hash in this case "photoalbum.eth". To discover the hash the ENS address is resolved to an account the contract held on the account can be searched for the "setContent" function as shown in Fig 6.



```
<!DOCTYPE html>
<!-- saved from url=(0047)https://swarm-gateways.net/bzz:/photoalbum.eth/ -->
<html slick-uniqueid="19"><head><meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
```

Figure 5. Web link that shows an ENS address in the HTML link from the page



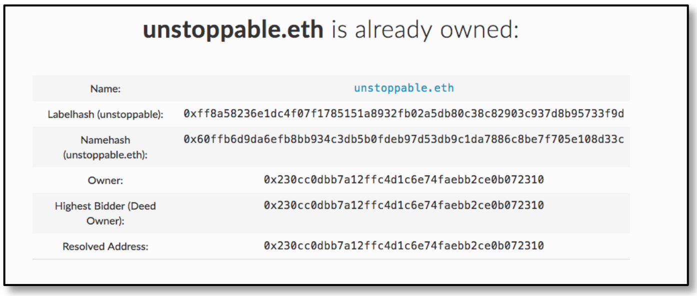
```
Function: setContent(bytes32 node, bytes32 hash)

MethodID: 0xc3d014d6
```

Figure 6. Method setContent function applying the hash to the node address

ENS names can also be applied to accounts so unstoppable.eth can applied to an Ethereum account / wallet and be used instead of the long account address. Fig 7 shows the name, resolver and account details assigned and revealed with in an ENS search within myetherwallet.com (Ropsten).

13. Discussion



unstoppable.eth is already owned:

Name:	unstoppable.eth
Labelhash (unstoppable):	0xff8a58236e1dc4f07f1785151a8932fb02a5db80c38c82903c937d8b95733f9d
Namehash (unstoppable.eth):	0x60ffb6d9da6efb8bb934c3db5b0fdeb97d53db9c1da7886c8be7f705e108d33c
Owner:	0x230cc0dbb7a12ffc4d1c6e74faebb2ce0b072310
Highest Bidder (Deed Owner):	0x230cc0dbb7a12ffc4d1c6e74faebb2ce0b072310
Resolved Address:	0x230cc0dbb7a12ffc4d1c6e74faebb2ce0b072310

Figure 7. ENS reverse lookup that shows the Name and additional bidding and resolved address

The ability to host decentralised resources and store material that would be traditionally held on centralised services changes some of the traditional methods of search. This scenario has demonstrated how material can persist beyond the normal experience of investigators creating an unstoppable hosting problem. The practical element demonstrated the use of the Ethereum network, just one of the technologies available to perform distributed storage. The ENS Ethereum Naming Service also provides the ability to link to TLD domains such as .xyz and .luxe. It is understood that the DNSSEC and the TLD integration will not likely resolve correctly with ENS as the DNS browser protocol may override the resolving [25]. There were a number of limitations and technical issues that could not be overcome to test a .xyz domain with any objectivity or confidence. The Ropsten testnet had some service issues during my testing with ENS and syncing, this included using the third-party API Infura that demonstrated the same behaviour. Where required I have used Ropsten and confirmed behaviour across the Mainnet with alternatively hosted sites. There are interesting uses of ENS and DNS hosted on Ethereum, the EthDNS system is prototyped and documented that uses DNS records stored on Ethereum [26]. There are potentially interesting attack vectors if Swarm and ENS became mainstream the use of a bad resolvers in new "Dapps" for example. IPFS also needs to be investigated to understand how it can be used in addition to existing technology or integration with other blockchain technologies. As the example shows the ability to bring up a node write information into the distributed storage is possible both quickly and cheaply, removal of the node from the system still allows the new files to remain. Attribution using a blockchain explorer allows account identification additional resources, identifiable information and linked smart contracts. The layers of investigation cut across web technology, blockchain account records, smart contracts, blockchain naming service, blockchain storage and the host machine. These can lead directly to additional accounts that may identify cryptocurrencies entering or leaving the system. The ability to interact with a smart contract using privacy focused technologies such as zkSNARKS or private smart contracts such as Enigma allow data or image sharing autonomously with strong encryption [27]. The ability to create a photo-sharing application for payment with content hosted on decentralised storage can be achieved using privacy focused methods in addition to blockchain technologies.

What is demonstrated is a need to understand the sources of hosted material as distributed storage becomes wider spread in its adoption. Hosting malware on distributed storage or indecent images of children will require investigators and responders to locate all the sources of material. In the examples shown it is possible to make attribution to file access and use for forensic examination. File signatures, hash values, hosted distributed domains, protocol specific URLs, e.g. BZZ or IPFS can be extracted. In incident response scenarios, the ability to source and collect the sample for reverse engineering will be essential for mitigation and research. Virus scanning and network protection rules could be used to search detect and block hosted material entering or leaving a network. Fig 8 shows the host and file access to the blockchain via Blockchain node / software or via internet gateways.

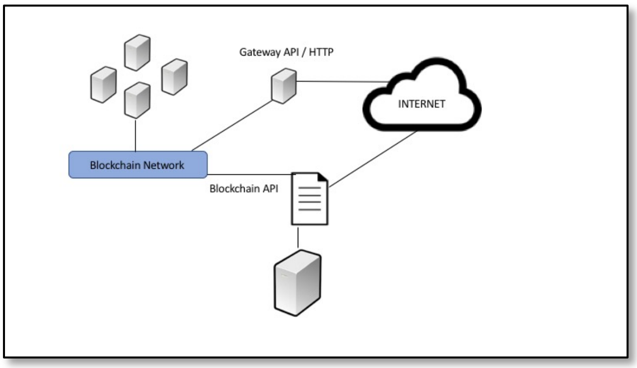


Fig 8 shows a host connecting via the blockchain protocols or via an internet gateway

14. Conclusion

This scenario has demonstrated it is possible to store content persistently on blockchain technology allowing access to those on the blockchain and to the internet through internet gateways. Decentralised storage remains uncensorable with no technical recourse to remove or even request for lawful motions against its storage. There are no regulations such as GDPR, local laws, state, or international law that have any power to control or remove it. The hosting of resources such as images or files on distributed file storage requires additional investigative methods to discover the source and linked information. The ability to attribute the access or presence of an illegal image or document can be reliable proven using hashing protocols used in Ethereum Swarm, the Swarm hash and the temporal data from the blockchain against fragments held on the host. The requirement to recover electronic data stored or what was accessed is needed in E-discovery and for corporate legal compliance, so the need exits to be able to seek and find documents hosted as described. Malware researchers require the source file to reverse engineer or perform static analysis so the ability to access blockchain storage to recover such files along with additional threat intelligence from linked accounts and blockchain naming is essential. In this case forensic artefacts were not interfered with in terms of their integrity, this is good news for forensic investigators wanting to review rich sources of meta-data. This was only performed on the Ethereum Swarm and other storage systems may also leave metadata or artefacts, a potentially important forensic research area. Research on distributed storage is still focused on the introduction, development, scalability and the performance of the technology. There are clearly vast gaps in literature around the use and long-term performance behaviour as the technology is rapidly evolving. Blockchain forensics has focused on cryptocurrency track and trace but the evolvement of smart contracts and now storage and computational resource will be a future frontier. It is unclear on the adoption of these technologies to long-term adoption, but a new challenge and knowledge gap could appear overnight. Blockchain will undoubtable continue to pioneer computational breakthroughs but new paradigms and challenges exist in its wake. The misuse cases should be considered and researched to compliment blockchain development as a global revolution.

Competing Interests:
None declared.

Ethical approval:
Not applicable.

Author's contribution:
SD has prepared the manuscript in entirety.

Funding:
None declared.

Acknowledgements:
None declared.

References:

[1] F. January, M. Ii, F. I. Directive, R. R. Review, E. Union, G. Data, P. Regulation, and T. Gdpr, "The rise of the regulator may lead to trouble for the blockchain," pp. 1–2, 2018.

[2] C. Salmensuu, "General Data Protection Regulation and the Blockchains," *Liikejuridiikka*, no. 1, p. 92, 2018.

[3] B. Ramsundar, R. Chen, A. Vasudev, R. Robbins, and A. Gorokh, "Tokenized Data Markets," 2018.

[4] N. Vergauwen, "Upgradeable Smart Contracts – Hacker Noon," *Medium - Hackernoon*, 2018. [Online]. Available: https://hackernoon.com/upgradeable-smart-contracts-a7e9aef76fdd. [Accessed: 15-Jan-2019].

[5] "Filecoin - Website," 2019. [Online]. Available: https://filecoin.io/. [Accessed: 15-Jan-2019].

[6] "IPFS is the Distributed Web," 2019. [Online]. Available: https://ipfs.io/. [Accessed: 15-Jan-2019].

[7] J. Redman, "BCH-Powered Bitcoin Files Project Adds IPFS Support - Bitcoin News," *News-Bitcoin.com*, 2018. [Online]. Available: https://news.bitcoin.com/bch-powered-bitcoin-files-project-adds-ipfs-support/. [Accessed: 06-Jan-2019].

[8] M. Zalecki, "Using IPFS with Ethereum for Data Storage | Tooplox," *TOOPLOOX - WEB*, 2018. [Online]. Available: https://www.tooplox.com/blog/using-ipfs-with-ethereum-for-data-storage. [Accessed: 06-Jan-2019].

[9] "MaidSafe," 2019. [Online]. Available: https://maidsafe.net/. [Accessed: 15-Jan-2019].

[10] "Sia," 2019. [Online]. Available: https://sia.tech/. [Accessed: 15-Jan-2019].

[11] "Storj - Decentralized Cloud Storage," *Storj - Decentralized Cloud Storage*. 15-Nov-2017.

[12] "Go Ethereum," *Geth*. [Online]. Available: https://geth.ethereum.org/. [Accessed: 15-Jan-2019].

[13] "1. Introduction — Swarm 0.3 documentation," *Swarm read the docs*, 2019. [Online]. Available: https://swarm-guide.readthedocs.io/en/latest/introduction.html. [Accessed: 15-Jan-2019].

[14] R. Amado, "How Cybercriminals are using Blockchain DNS | Digital Shadows," *Digital Shadows_ (Web)*, 2018. [Online]. Available: https://www.digitalshadows.com/blog-and-research/how-cybercriminals-are-using-blockchain-dns-from-the-market-to-the-bazar/. [Accessed: 14-Jan-2019].

[15] "Namecoin," *Namecoin (Web)*, 2019. [Online]. Available: https://namecoin.org/. [Accessed: 14-Jan-2019].

[16] M. Ali, J. Nelson, R. Shea, and M. J. Freedman, "Blockstack : A Global Naming and Storage System Secured by Blockchains," *USENIX Annu. Tech. Conf.*, pp. 181–194, 2016.

[17] I. Ilascu, "New Botnet Hides in Blockchain DNS Mist and Removes Cryptominer," *Bleeping Computer - Web*, 2018. [Online]. Available: https://www.bleepingcomputer.com/news/security/new-botnet-hides-in-blockchain-dns-mist-and-removes-cryptominer/. [Accessed: 14-Jan-2019].

[18] N. Johnson, "A developer's guide to ENS concepts – The Ethereum Name Service – Medium," *Medium Blogpost Web*, 2017. [Online]. Available: https://medium.com/the-ethereum-name-service/a-developers-guide-to-ens-concepts-7004eea8a073. [Accessed: 13-Jan-2019].

[19] M. Bazzell, *Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information*, 5th ed. USA: CreateSpace Independent Publishing Platform, 2016.

[20] D. Murdoch, *Blue Team Handbook: Incident Response Edition*. 2014.

[21] "Personal data breaches," 2019.

[22] R. Jones and P. Collinson, "Identity theft warning after major data breach at Ticketmaster | Money | The Guardian," *The Guardian (Online)*, 2018. [Online]. Available: https://www.theguardian.com/money/2018/jun/27/identity-theft-warning-after-major-data-breach-at-ticketmaster. [Accessed: 06-Jan-2019].

[23] D. P. Act, "ICO lo Guidance on data security breach management," pp. 1–8, 1998.

[24] The Crown, "PML v Person(s) Unknown [2018] EWHC 838 (QB) (17 April 2018)," 2018.

[25] N. Johnson, "ethereum/go-ethereum/name-registry - Gitter," *Chatboard*, 2019. [Online]. Available: https://gitter.im/ethereum/go-ethereum/name-registry. [Accessed: 15-Jan-2019].

[26] J. McDonald, "EthDNS: an Ethereum backend for the Domain Name System," *Medium Blogpost Web*, 2018. [Online]. Available: https://medium.com/@jgm.orinoco/ethdns-an-ethereum-backend-for-the-domain-name-system-d52dabd904b3. [Accessed: 13-Jan-2019].

[27] S. Dyson, W. J. Buchanan, and L. Bell, "The Challenges of Investigating Cryptocurrencies and Blockchain Related Crime," vol. 1, no. 2, pp. 1–6, 2018.