The SAFE Network
a New, Decentralised Internet

Nick Lambert[*] and Benjamin Bollen[†]

October 2-3, 2014
Paper for the final symposium of the ADAM project

Abstract

The Internet is an incredible resource, enabling the storage and sharing of data amongst 40% of the world's population. However, these storage locations are inherently insecure and enable mass surveillance and data theft by companies and world Governments. This paper proposes a solution by redesigning and reimplementing the Internet's underlying infrastructure to require no central control and by implication, no servers as we currently know them. The ideas presented here allow the creation of a network, which provides all users the opportunity to retain complete control of their own security and personal information.

1 Introduction

Computing capability has increased dramatically in recent years, enabling the creation of some incredible advancements. The distribution of a faster Internet has also played its part with an estimated rise in average worldwide connection (download) speed from 2.1Mbps in Q1 2011 to 3.9 Mbps in Q1 2014[1]. These technological advancements, faster infrastructure and greater availability have led to products and services that were previously not possible. For example, the delivery of voice and multimedia communications (VOIP) between mobile devices are now common place[1]. These advancements have not only provided the human race with a greater suite of tools than ever before, they have also led to fundamental changes in how we communicate with each other. The rise of social networking websites, mobile devices and mobile applications has led to an explosion in what and how much data we share with each other.

This technological evolution is not only affecting the volume of information sharing, it is also heavily impacting upon where consumers store data. ABI research reported the existence of 1 billion personal cloud storage accounts in 2013 with an average 685 MB stored per account. The research forecasts a growth to 3.61 billion accounts by 2018 with an average of 975 MB per account[4]. These storage providers can be either ecosystem companies, or core cloud storage companies[2], but crucially in both cases, a private company owns and controls the data, not the end-user.

So while these enhancements take technological ease of use to the next level, these convenient and often free services come at a price. The price is freedom. Today's Internet architecture, where centralised and managed intermediaries (servers) store

[*] nick.lambert@maidsafe.net

[†] benjamin.bollen@maidsafe.net

[1] For example, Skype usage in 2013 accounted for 214 billion minutes in 2013[2], or roughly 400.000 active connections on average. Since 2005 VOIP is increasingly dominating the growth of long-distance communication at the expense of conventional phone calls[3]

[2] Industry leaders for ecosystem providers are for example Apple, Google, Microsoft, Alibaba or Yandex. For core cloud storage companies the landscape is more dynamic, but Dropbox for personal storage or Box for business collaboration are prominent names.

and provide access to information, do so in an inherently insecure way. We also experience an ever increasing number of concerns over the privacy of our data. This paper argues that it is in fact human involvement, and the existing client server architecture's requirement for human organisation, that leads to these security and privacy issues.

MaidSafe.net, a Scottish company founded for these goals, proposes and implements a new, decentralised architecture that eliminates human involvement in private data. MaidSafe will enable users to enjoy the significant resources of the Internet, while improving and protecting their experience and privacy.

2 The Internet is broken

The fact that the Internet has grown beyond the expected use cases of the original design is, at the very least, a strong motivation to consider a renewed architecture. It is evident looking back that the current volume of 2.8 billion regular users[5] was not anticipated, nor was the original design of ARPANET centralising. In fact, one of Bob Kahn's fundamental rules, when designing the transmission control protocol (TCP), was that there would be no global control at the operations level[6]. However, some of these principles took a back seat as other considerations took priority.

It was originally envisioned, back in the late 1960s, that there would be multiple independent networks and as Leiner et al suggested "256 networks would be sufficient for the foreseeable future". This was clearly in need of consideration when Local Area Networks (LANS) began to appear in the late 1970s. The addition of workstations, PCs and Ethernet technology, in addition to LANs, also led to changes in the original architecture concepts. The rapid and unforeseen rise in the Internet's growth introduced scaling issues[3] that were dealt with by the implementation of a hierarchical routing model. This approach led to a centralising of the architecture, with the introduction of "managed interconnection points" by US Federal agencies.

This enabled more "rapid configuration robustness and better scaling to be accommodated". As the National Science Foundation (NSF) started to privatise and commercialise the program in 1995, the use of regional networks via private long haul carriers led to the information superhighway. This made the world wide web, envisioned by Tim Berners-Lee, possible. However, as the Internet has continued to grow, it is suggested that this change in direction has led to some significant problems that not only impact upon the way the world's citizens manage data, it is also having a much more profound impact on society as a whole.

2.1 Government Control

Governments spying and eavesdropping is nothing new[7]. Letters have been intercepted for centuries and spy networks have existed for millennia. In recent times, attention has shifted from telegrams to email. Within the US, Project Shamrock, established after the Second World War, legally accessed all the cables of RCA Global, ITT, and Western Union. Minaret was another project, established in the 1960s to focus on intelligence gathering amongst domestic targets. Despite both programs being exposed and subsequently shelved, they were the pre-cursor to ECHELON, a surveillance program that utilises automatic keyword searching of faxes, telex and emails. In 2001 the European Parliament examined the reach of the ECHELON program, but concluded the impact would be limited[8]:

> [. . .] however extensive the resources and capabilities for the interception of communications may be, the extremely high volume of traffic makes exhaustive, detailed monitoring of all communications impossible in practice.

Technological advancement over the past decade has refuted the assumption of this report. Centralisation driven by cloud services has facilitated the impact of surveillance programs. The Snowden revelations demonstrated that intelligence agen-

[3] The single algorithm employed with all routers could not cope with demand.

cies are able to access data at will from some of the world's largest technology companies (PRISM) and by tapping data direct from fibre optic cables (TEMPORA, FAIRVIEW, STORMBREW, OAKSTAR and BLARNEY)[9]. Although denied by almost all technology companies, NSA slides suggest they were complicit, willing or otherwise, in helping to collect this data[10].

This paper proposes that the existing centralised architecture and the involvement of humans, make these privacy intrusions possible. All existing services require that users authenticate themselves in order to gain access. While this process is automated, the credentials of the users are stored in centralised locations. Furthermore, the encryption keys are held by the service provider, enabling them to access their users' data[11]. While client side encryption technology is available, it is not standard practice to implement it in end-user products. Additionally, most cloud service companies are based within the US and as such are legally obliged to comply with Government agency requests for user data.

Interestingly, much of this monitoring conflicts with the Universal Declaration of Human Rights that the US and UK Government and their allies signed up to in 1948. Article 12 states[12]

> "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."

The evidence suggests that, over a lengthy period, these Governments have chosen to ignore a fundamental human right, the individual's right to privacy. However, it is not just nations that choose to invade our privacy. As the Internet was designed without inherent privacy protections at protocol level, technology companies also take advantage of the centralised Internet architecture to profit from private data.

2.2 Surveillance as a business model

Security expert Bruce Schneier notably stated at SOURCE conference in Boston:[13]

> "Surveillance is the business model of the Internet. We build systems that spy on people in exchange for services. Corporations call it marketing."

This suggestion is based on the fact that many large technology companies generate the overwhelming majority of their revenue by mining their users' data. The revenue model of companies like Google and Facebook is to provide their core services free from monetary charge and then sell user-targeted access to companies who advertise on their platform. Google generated USD 50.5 billion in advertising revenue during 2013 which equated to 91% of their yearly sales[14]. Similarly, Facebook delivered advertising revenue of over USD 6.9 billion to their investors over the same period, equivalent to 89% of their income[15].

Ethan Zuckerman (Director of the Center for Civic Media at MIT and principal research scientist at MIT's Media Lab) has suggested that[16]:

> "The fallen state of our Internet is a direct, if unintentional, consequence of choosing advertising as the default model to support online content and services."

Zuckerman goes on to argue that Facebook, Google and others, are under increasing pressure from shareholders to sell ever more advertisements. They predominantly grow their advertisement revenues by providing ever deeper insights into their users. As a result, such companies mine their users' data at an ever increasing rate.

This paper suggests that the ad supported web, while not being directly responsible for the move toward a more centralised architecture, has nonetheless had a part to play. It is much easier to mine our data when it is held in unencrypted and centralised locations under direct control of the service provider.

2.3 Security

Security of data is an issue that goes hand in hand with privacy, as without security data cannot remain private. A European Commission project, the

ABC4Trust, suggests that users want privacy and organisations want security, however, it could be argued that information stolen from a business typically also affects individuals and therefore both groups desire security, albeit of differing importance[17].

Whilst opinion on this point may vary from reader to reader, evidence suggests that security is about as scarce as privacy within today's centralised Internet. At a time when the revenue generated by the Internet as a whole is estimated to exceed USD 4.2 trillion in G20 countries alone, and with companies storing and sharing highly valued and sensitive information, demand for Internet data security has never been greater[18].

In a UK Government survey on security breaches, 93% of large organisations and 87% of SMEs experienced a security breach in 2013. The survey reported that major breaches were typically caused by a combination of failures in not only people, but also process and the technology itself[19]:

> "36% of the worst security breaches in the year were caused by inadvertent human error (and a further 10% by deliberate misuse of systems by staff)"

What is also evident is that these problems are not sector specific and are taking place across all industries including healthcare, finance and education[20]

With global cyber crime costing an estimated USD 400 billion per year, it has a huge impact on society, companies and consumers[21]. In December 2013, 40 million credit card details were stolen from a prominent US retailer with the haul also including 70 million addresses and other personal information. Similarly, in what appears to be a series of attacks, a Russian group has amassed an estimated 1.2 billion user names and passwords[22].

With businesses continuing to store valuable information in centralised, hence inherently insecure, data centres, managed by people that empirical evidence indicates are prone to mistakes, suggests a change in approach is required. Failure to adapt to these issues will only ensure that they continue. What is required is a change to the fundamental architecture of the Internet; one that removes central points of weakness and humans from the process of data management. This fresh approach should clearly have security and privacy inherent within the design.

3 The SAFE network; a new network design

Taking inspiration from complex natural systems, and specifically ant colonies, Scottish engineer David Irvine set about the problem of providing a secure data and communications platform. MaidSafe.net was founded in 2006[4] to implement his design. The values of the company are to provide privacy, security and freedom for all. The SAFE network is open source and provided free of charge to the world.

The system is a fully decentralised, serverless, peer-to-peer design with the objective of providing an autonomous global network. The network is comprised by the users, who each donate their spare computing resources to it and are incentivised via a network token for doing so. Each user establishes a node, called a vault, to which the network assigns its own unique address (derived from a cryptographic key pair). This address is known only by the network.[23]

It is possible to store any type of data (structured and unstructured) on the vault network, with typically each user running a client to enable network requests. Vaults on the network don't only store data, they perform multiple functions, called personas. This includes managing the integrity of close nodes[5] and the integrity of the data chunks.

The closest nodes will continually change as nodes go on and offline. This provides a dynamic environment. Normal usage should induce a high network churn. Knowledge of a network area is managed by the concept of close groups. A close group are the vaults whose addresses are closest to a given address, but not equal to it. This determines a close group for every data chunk, but equally a close group is defined for every vault (figure 1). A majority-based decision algorithm ensures that each node follows the

[4] MaidSafe is an acronym: Massive Array of Internet Disks, Secure Access For Everyone.

[5] Distance is measured with a bitwise exclusive or operation on the network addresses.

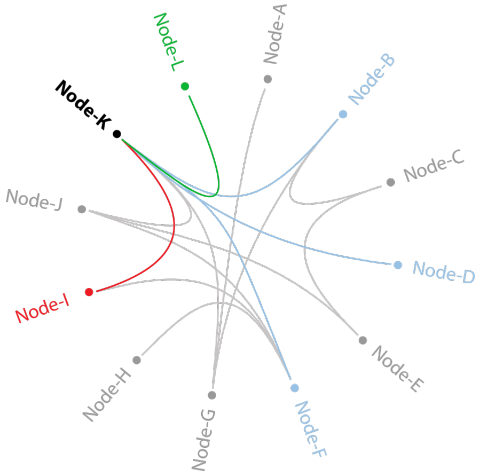

Figure 1: Illustrative connection map of nodes showing the dynamic nature of the network. Connected lines indicate that these nodes have each other in their routing tables. Nodes B, D, F and L are assumed to be the new close group for Node-K, where Node-I is no longer in the close group of four.

rules of the network. It minimises undesirable or malicious behaviour, as verifiable wrong calls by a node will be monitored by its surrounding nodes. These nodes will reduce the rank of the offending node and eventually exclude it from the network.[25]

Data is stored on the network in encrypted chunks. These chunks are generated by the Self-Encryption process where a sliding window encrypts and obfuscates chunks with the hash of neighbouring chunks. The final hash of the chunk serves as the name of the chunk. This makes the network content addressable, at least when you have access to the data map to reconstruct a file[26]. Additionally, this allows the network to self-heal stored data; a stored chunk whose hash does not correspond to its address has been corrupted and can be restored from redundant copies.

In the interest of data availability, the network retains minimally four live copies of each data chunk at any time. Data managers are responsible for ensuring a new copy gets created as vaults go offline. They also maintain a record of dead copies, as offline vaults are likely to return online later. Because the network is content addressable, deduplication of data is utilised.

An internal study in partnership with NHS Ayrshire and Arran in 2011, showed that MaidSafe's algorithm achieved a 48% deduplication rate on a 397GB data set, 312GB of which were DICOM x-ray images.

Importantly though, the MaidSafe network is the first distributed hash storage system that supports deletion of data. The algorithm accomplishes this without explicitly listing the registered owners of chunks in the network.

3.1 Data in a hostile environment

The problems of data surveillance, security and privacy are multi-faceted. The SAFE network employs a range of strategies to address these problems. By decentralising the storage of data on a global network, the SAFE network makes it more challenging for any local actor to monitor and trace information. Data is no longer confined to a physical storage location. Network churn makes storage dynamic and mostly a memory-to-memory operation. As data chunks are stored with a 512bit network address and each chunk of data is not linked to other chunks from the same file, reconstruction of a file without a datamap is computationally infeasible.

Anonymity is a crucial element of the network. To provide a fully decentralised network, users must be able to self-authenticate onto it. This process does not involve any register of the active users, or a third-party verification step. Users are able to generate their own credentials and store their identity securely on the network. The credentials allow a user to retrieve a passport from the network that stores all additional keys to access and reconstruct their stored content from the network.[27]. This provides users with the ability to manage their own identity for the first time at network level. The authentication process happens fully within the client computer. The password is never sent out onto the network.

Self-Authentication within the SAFE network provides users with anonymity as no entity, MaidSafe included, is aware of any information about any of the network users. Self-Authentication also ensures that access to the network can never be restricted for any individual.

This means that applications on top of the SAFE network need to get explicit permission from any user to collect personal information. The user can always retract these permissions and is the owner of his/her own data at all times.

The Self-Encryption process of all data ensures that data on the network is always encrypted, whether at rest or in transit. Data is only ever decrypted by the client, on the end-user machine. The encryption password cannot be stolen from the network, as it is not given out.

However, the risk remains that credentials are stolen from a compromised end-user machine. There are measures one can take to protect against this threat, like external hardware authentication keys. Such measures are currently beyond the scope of this article and the MaidSafe project. As keylogging and hacking into individual end-user machines is a resource intensive tasks for any organisation, the SAFE network levels the playing field for the opposing forces of a right to privacy and a drive for mass surveillance.

It is also worth considering the robustness that the SAFE network provides. As the network is comprised of the resources of its users, as opposed to a central location, it cannot be turned off and no kill switch exists. Furthermore, the network does not use the Domain Name System (DNS), making it impervious to web censoring.

All SAFE traffic exists as fully encrypted UDP packets. This implements Net Neutrality at the core of the SAFE network. All data packets are indistinguishable and can only be treated equally.

However, despite many of these advantages, there are many challenges the network needs to overcome to achieve widescale adoption.

4 Discussion

4.1 Adoption challenges

It is proposed that the SAFE network offers a solution to many of the security and privacy challenges experienced with the Internet today. However, even if the implementation and widescale adoption of the network is assumed, it still leaves many considerations for which answers are not yet known. Governments will still want to surveil both their citizens and others whom they deem to be a potential threat.

As mentioned before, credentials can be stolen and human beings are often the weakest link in a security scheme. This paper proposes that Governments may have a legitimate interest to surveil possible security threats. However, with the introduction of the SAFE network, mass surveillance will become too resource intensive to maintain.

Attacks can also take a non-technical form. For example, public relations efforts to discredit the network to the public, slowing and even halting adoption are a possibility.

Removing advertising as a default form of payment for online services will also require significant adjustment and many companies who experience success with the status quo might be resistant to change.

However, it is important that the SAFE network does not make the advertisement driven business model impossible. On the contrary, the SAFE network drastically cuts the infrastructure costs of online services, and a service may allow users to actively choose to pay for their usage by receiving advertisements. The SAFE network just restores the choice to the users.

Alternatively, cryptocurrencies can be part of the solution. Innovations such as Bitcoin (currently) provide very low transaction fees and can be divided with a resolution of 10^{-8} BTC, making micropayments a viable option. Accumulated micropayments can automatically be transferred to the correct rights holders, be it for text, music, movies or applications. At present, Bitcoin has some technical challenges regarding transaction speed[6] that could limit its usefulness. Also the real transaction cost - disregarding mining - for a bitcoin transfer is not insignificant. This cost is volatile, but has averaged above USD 20 per transaction since December 2013[28].

Interestingly, the SAFE network incorporates a new cryptocurrency technology, called Safecoin. Safecoin serves first to incentivise constructive behaviour on the network. Users are rewarded with Safecoin, paid to them by the network, as compensation for providing their computing resources. Similarly, application developers are rewarded for creating applications, based on how much their applications are used. As Safecoin transactions are not chained[7] and a more efficient method of consensus is used, ie close groups within the network, it is energy-efficient and it can validate transactions at network speed. Additionally, Safecoin guarantees micropayments with a divisibility of at least 2^{-32}. These attributes mean that it is a potentially viable way of facilitating rapid micropayments for any type of content within the network.

Clearly there is no obvious answer to the question of how to fund a new decentralised Internet, and it is out the scope of this paper to attempt to find one. However, there are many alternatives, all of which put the user in control of their data.

4.2 Competing alternatives

MaidSafe is not the only organisation to build decentralised technologies for network infrastructure. There are a number of alternative projects that value decentralisation. It is outwith the scope of this paper to provide a detailed analysis of each. The intention is to acknowledge that the SAFE network is not an isolated project.

It should be noted, however, that these listed alternatives are restricted to public publishing networks only for the sake of brevity. The reach of the MaidSafe protocol is more fundamental and supports any type of Internet or private application. Additionally, the simple principles of its operation allow for extension to new functionalities in the future with a particular interest in decentralised computation.

Established in 2000, Gnutella was one of the earliest decentralised pure peer-to-peer networks and currently supports several million users. As with the SAFE network there is no reliance on any central servers. The network is comprised of leaf nodes, nodes which have no child nodes, and ultra nodes, which are capable of routing requests and responses from other nodes on the network. This is accomplished via the exchange of a Query Routing Table. Peers communicate using an application level protocol which uses message headers that describe the message payload. Peers send the messages they receive to other peers they are connected to. In this respect, messages flood the network, making the process less efficient than some of the other alternatives.

The Gnutella network has been successfully utilised by clients, such as Emule, which enable public file sharing. However, the implementation does not permit private data to be stored. Furthermore, data on the Gnutella network is also intended to be immutable and therefore does not enable data removal. Whilst these attributes make it well suited to public file sharing services, it would not be a suitable replacement for all existing web services.

Freenet is another peer-to-peer network that utilises a decentralised data store that provides its users with anonymity protection and censor-resistant communications. The open source project was established in 1999 and his been in development ever since.

[6] It can take up to 60 minutes to confirm a Bitcoin transaction from several miners.

[7] Only the previous and current owner of the coin is known.

The network is comprised of multiple nodes, some of which act as hosts for data and others which only route the flow of data. Every node on the network contributes storage space and data is stored in encrypted blocks and spread on several network nodes. Unlike the SAFE network, which was designed as a decentralised data and communications network, Freenet was primarily designed as a network for the publication of files and is not intended to limit access to those files. There is no mechanism to delete data on the network, however, information that is not retrieved regularly can drop off the network as allocated disk space is utilised. This approach makes Freenet an effective anonymous publishing platform, but its lack of storage and management of private data limits its potential to replace today's existing centralised web services.

BitTorrent is the most popular peer-to-peer network and is, according to the company, used by an estimated 150 million users world wide. BitTorrent brings an innovative approach to the problem of scalability within P2P systems, which typically rely on source peers to provide the majority of the resource when downloading large files. To improve efficiency, BitTorrent splits each file into 256kB blocks and these are passed between peers on a quid pro quo basis. File fragments can then be downloaded out of sequence, enabling peers to obtain fragments they require from other peers on the network, thus providing faster downloads and a fairer distribution of resources between network peers.

However, BitTorrent is not a completely decentralised network, utilising servers (torrent trackers) to monitor where file blocks are located in order to assist in the transmission and reconstitution of the entire file. Clients require to communicate with the tracker servers in order to initiate downloads. The use of these centralised points not only represents a potential point of weakness, they are also a potential security risk as they represent a public directory for the location of file blocks. Increasingly though, the original torrent files have been replaced since several years with a hash based universal resource identifier (URI), a so called magnetic link. This makes the BitTorrent network much more decentralised, while it still relies on web searches for discovering the magnetic URI for specific content.

Conclusion

The SAFE network potentially provides a solution to those looking to enjoy the vast resources of the Internet without many of the downsides, which include mass surveillance from governments and companies. The SAFE network also aims to minimise many of the security risks that currently exist with the existing World Wide Web. The SAFE network has been implemented in a decentralised architecture and has been designed in this way to remove the requirement for human intervention from our data, while also removing servers, which act as a central point of weakness.

References

[1] Statista, *Average Global Internet Connection Speed from 1st Quarter 2011 to 1st Quarter 2014*, statista.com 2014

[2] J.R. Nay, *Skype's Worldwide Traffic Continues to Grow, with 214 Billion Minutes on VOIP in 2013*, TruTower, January 13, 2014

[3] PriMetrica, *TeleGeography Report 2013*, Executive Summary

[4] ABI Cloud Content and Services Research service, *Personal Cloud Storage and Synchronization Study*, Personal Cloud Storage Accounts Total One Billion in 2013, Generating 685 Petabytes, Research News December 13, 2013

[5] Internet World Stats, *Internet Users in the World Q4 2013*, Miniwatts Marketing Group Internet World Stats

[6] V. Cerf et al, *Brief History of the Internet*, Internet Society

[7] A. Zurcher, *Roman Empire to the NSA: A World History of Government Spying*, BBC News, November 1, 2013

[8] G. Schmid, *On the existence of a global system for the interception of private and commercial communications (ECHELON interception system)*, European Parliament report PE 305.391, A5-0264/2001, July 2001

[9] The Guardian, *The NSA Files*, www.theguardian.com/world/the-nsa-files

[10] G. Greenwald and E. MacASkill, *NSA Prism program taps in to user data of Apple, Google and others*, The Guardian, June 7 2013

[11] P. Lambert, *Does your cloud storage provider hold the keys to your data*, TechRepublic.com, April 9, 2012

[12] United Nations General Assembly, *The Universal Declaration of Human Rights*, Human Rights Declaration, article 12, December 10, 1948

[13] SOURCE conference Boston 14, *keynote speech* by Bruce Schneier, Boston, April 2014; F.Y. Rashid *Surveillance is the Business Model of the Internet: Bruce Schneier*, Security Week April 9, 2014

[14] Google Investor Relations, *Google's Income Statement Information*, Financial Tables 2013

[15] Facebook Investor Relations, *Facebook reports Fourth Quarter and Full Year 2013 Results*, Fourth Quarter and Full Year 2013 Financial Summary

[16] E. Zuckerman, *The Internet's Original Sin*, The Atlantic, August 14, 2014

[17] European Commission, *Security and privacy? Now they can go hand in hand*, Digital Agenda for Europe, News, May 27, 2014

[18] The Boston Consulting Group, *The Internet Economy in the G-20*, BCG report, the connected world, the internet economy in the G-20, March 2012

[19] Department for Business Innovation and Skills, UK government, *2013 Information Security Breaches Survey*, Executive Summary BIS/13/P184 - ES, technical report 2013

[20] Symantec, *In Defense of Data*, Symantec Official Blog, January 1, 2014

[21] McAfee, *Stopping Cybercrime Can Positively Impact World Economies*, McAfee Press Release June 9, 2014

[22] N. Perlroth and D. Gelles, *Russian Hackers Amass Over a Billion Internet Passwords*, New York Times August 5, 2014

[23] G. Paul, F. Hutchison, J.Irvine, *Security of the MaidSafe Vault Network*, in collaboration with University of Strathclyde

[24] MaidSafe, *Core Development Introduction*, maidsafe.net/core-developers

[25] MaidSafe, *Vault Documentation*, GitHub MaidSafe-Vault documentation

[26] MaidSafe, *MaidSafe Encrypt Library*, MaidSafe-Encrypt overview

[27] MaidSafe, *MaidSafe Self-Authentication*, MaidSafe Self-Authentication paper, September 2010

[28] Blockchain, *Cost per Transaction*, chart produced by blockchain.info, time of writing September 2014