Security Recommendations for organisations employing a Hybrid Cloud model

A dissertation submitted in partial fulfilment of the requirements for the degree of Bachelor of Science (Honours) in Software Engineering

by S. D. Bean

Department of Computing & Information Systems
Cardiff School of Management

Cardiff Metropolitan University

April 2017
Declaration

I hereby declare that this dissertation entitled *Security Recommendations for organisations employing a Hybrid Cloud model* is entirely my own work, and it has never been submitted nor is it currently being submitted for any other degree.

Candidate:  *Samuel Bean*

Signature:

Date:

Supervisor:  *Paul Angel*

Signature:

Date:
“Security is not a product, but a process.”

- Bruce Schneier
Contents

1.0 Abstract  p. v
2.0 Acknowledgments  p. vi
3.0 List of Figures  p. vii
4.0 Introduction  p. 2
  4.1 Network and Application attacks  p. 4
  4.2 Encryption  p. 5
  4.3 Virtualisation and Multi-Tenancy  p. 6
  4.4 Data Loss  p. 6
  4.5 Legality and Policies  p. 7
5.0 Literature Review  p. 10
  5.1 Network and Application attacks  p. 10
  5.1.1 Network and Application attacks Analysis  p. 12
  5.2 Encryption  p. 12
  5.2.1 Encryption Analysis  p. 14
  5.3 Virtualisation and Multi-Tenancy  p. 14
  5.3.1 Virtualisation and Multi-Tenancy Analysis  p. 16
  5.4 Data Loss  p. 16
  5.4.1 Data Loss Analysis  p. 18
  5.5 Legality and Policies  p. 18
  5.5.1 Legality and Policies Analysis  p. 20
6.0 Evaluation & Recommendations  p. 22
  6.1 Network and Application attacks  p. 22
  6.1.1 DDoS  p. 22
  6.1.2 Data Access Control  p. 24
  6.2 Encryption  p. 26
  6.3 Virtualisation and Multi-Tenancy  p. 29
  6.4 Data Loss  p. 31
  6.5 Legality and Policies  p. 34
7.0 Conclusion  p. 38
  7.1 Network and Application attacks  p. 38
7.2 Encryption  p. 39
7.3 Virtualisation and Multi-Tenancy  p. 40
7.4 Data Loss  p. 40
7.5 Legality & Policies  p. 41
7.6 Summary  p. 42

8.0 Further Research  p. 43
9.0 Appendix  p. 44
10.0 References  p. 51
1.0 Abstract

Modern business commonly sees the use of Cloud Computing to reduce costs and allow ease of use for mobile business. Cloud platforms offer an easy solution to lots of hardware and other hindrances to business operations. However, a more open and lightweight system means that security flaws will become apparent. While security is not as secure as it can be, organisations will find that productivity decreases meaning that their business is less efficient. Here, we will be addressing 5 areas where security threats are most likely to prevail. Using existing research and modern security countermeasures, recommendations will be provided for businesses and organisations to tighten security when a Hybrid Cloud model is employed. Recommendations made will aim to provide enough security to negate the majority of threats while still allowing day to day operations to be carried out as normal.
2.0 Acknowledgements

Paul Angel

Thank you for providing helpful insight into structure, layout and contents of this document and being constantly available when advice was needed. Thank you for being an overall great tutor.

My Friends & Family

I would like to thank you for all the encouragement and advice when writing my dissertation. For taking time to proof read my work and being supportive when stress levels were high. And for also providing guidance towards this project.

Cardiff Metropolitan University

Thank you for providing the necessary resources to allow me to construct the best piece of work possible. To the staff for providing me with the knowledge of my topic area, for without this my work would not have been possible.
3.0 List of Figures

**Figure 1:** Service management between Cloud Infrastructure, Platform and Service – Page 4

**Figure 2:** Security Defence in Depth model – Page 7

**Figure 3:** Optimal neural network representation using ANN – Page 23

**Figure 4:** Hybrid encryption model – Page 29

**Figure 5:** Distribution of hardware resources within virtualisation – Page 30

**Figure 6:** Probabilistic data checking representation – page 32
Introduction
4.0 Introduction

Cloud Computing has been an increasing asset to many businesses and members of the public over the last decade. It allows for up front hardware costs to be reduced, making it an attractive option for both start-up and well established businesses. Cloud services appear in 3 main forms: infrastructure, platform and software (Mell et al. 2011). But for everything good the cloud offers, there is also a posing security threat awaiting exploitation. Since data lacks a completely secure link between the provider and the client, data security is a significant liability for cloud users.

Many businesses have begun to make the conversion to a hybrid cloud model for a number of reasons. Having a split between a public and private network gives the user flexibility to control a portion of their own privately run cloud with flexibility as to what is done with data stored and processed on the cloud. Whilst having a public domain for storing data that is easily accessible from anywhere. 200,000 businesses are currently using DropBox’s business service and therefore 200,000 cloud environments being used by varying sizes of businesses every day (DropBox 2017). Private cloud statistics are harder to access however if we are to assume that there are a similar number of private Cloud Computing networks then 400,000 networks exist from a single CSP alone.

The main Cloud models most commonly adopted are: Public, Private and hybrid models. The public cloud offers a network open to be accessed by everyone and are often provided by big service providers such as Google or Microsoft. Since there are many users across the public cloud network, security is typically intense so that data is protected. Infrastructure can also be shared between several organisations to generate a larger infrastructure, also known as the Community Cloud. The organisations usually share security concerns, meaning they can adopt a bigger network with split costs of the service and more importantly, fewer security risks. Within a private network, internal security can be loosened since the users should be granted access to the system via an administrator. However, this doesn’t mean that there should be no security at all.
Private models are usually owned by organisations that have a lot of private data to store and manage, making them very desirable to hackers. If the private network has no connection to the outside world it is less of a risk, but when there is a link between a private network and a public network, this is when security becomes more challenging. The Hybrid cloud is a combination of both public and private networks, often taking advantages from both network models.

Infrastructure as a Service (IaaS) is a service type of Cloud computing that maintains a self-service model. This offers computer resources as well as storage hosted via service providers. IaaS is a self-provisional infrastructure, allowing its consumers to use a Web-based graphical user interface that serves as an IT operations management console for the overall environment (Gartner 2017).

Platform as a Service (PaaS) is mainly used for development and consumer application hosting. This offers an environment where cloud components can be implemented into user software. This allows full customisation and creativity to be offered to developers. A 3rd party to the service provider manages all aspects including operating system, virtualization, servers, storage, software and networking (Gartner 2017).

Software as a Service (SaaS) includes many web based applications that can be run with little to no installation necessary with the exception of plugins. SaaS is currently the biggest area of the Cloud market today. A typical SaaS offering would include email applications, maintained and managed by vendors (Gartner 2017).
Figure 1 shows the divide of public service management between the CSP and the consumer. Infrastructure offers more consumer freedom such as Security and applications used. However if more control is desired over attributes, a privately hosted Cloud should be used.

Fundamentally, data security is commonly defined under 3 main properties. Confidentiality is the assurance that sensitive data is protected against unauthorised personnel and software. Integrity is the assertion that data, programs and other system resources are protected against malicious activity or involuntary modification via unauthorised personnel. And finally, availability is assurance that the rightful owners of the data, will not be denied access to the data that is stored regarding or belonging to them (Reilly 2004).

4.1 Network and Application attacks

With attacks on networks and Cloud hosted software being a paramount threat in the CC industry, Cloud Service Providers (CSPs) and 3\textsuperscript{rd} party software developers must ensure necessary countermeasures to prevent and potentially
negate attacks. Network attacks are often detected in the form of Distributed Denial of Service (DDoS) attacks. Initiated by a Trojan injection, botnets infiltrate multiple machines, flooding the servers with requests. Distributed attacks mean administrators cannot simply block a single IP to prevent the attack. Infected traffic is also particularly hard to distinguish when compared to normal traffic due to the packets being forged (Daintith and Wright 2008). Blocking infected requests is a case of developing an algorithm that can distinguish between the two types of traffic. Algorithms are often implemented in firewall software or other types of Anti-DDoS software making them more accessible for users. We will be discussing several different algorithms used to fight DDoS attacks and analysing their benefits and drawbacks.

4.2 Encryption

Encryption is a vital aspect of maintaining data security. Different encryption methods hash data differently but the end product should mean parties without access are unable to construe anything from the data. In most instances, a secret key must be obtained to convert the cipher text into plain text (Daintith and Wright 2008). Early encryption algorithms started out with Data Encryption Standard (DES), which was soon cracked by deciphers with some ease. To counter this Triple DES (TDES) was initiated in 1998. TDES uses 3 individual 56 bit keys totalling 168 bits. TDES was used in financial services and other industry before slowly phasing out. RSA is one of the first public key cryptosystems. Unlike DES, it uses an asymmetric algorithm as it uses a pair of keys which are different, to decipher. To this current day Advanced Encryption Standard (AES) is the US government trusted encryption algorithm (Mell et al. 2011). To act as a primary layer of defence, many businesses encrypt their data to prevent it falling into the wrong hands. Several encryption methods will be analysed and the most efficient and effective method will be put forward to operate within a hybrid cloud environment. Good attributes for encryption algorithms to be used with a hybrid cloud include compatibility between parties, complexity of hashing and ease of use.
4.3 Virtualisation & Multi-Tenancy

A great attribute adopted by many hybrid SaaS models within businesses is called Virtualisation. This is defined as creating a replica of hardware without the hardware being present. We see this a lot with concepts such as virtual machines. A more specific area of virtualisation within the hybrid cloud is known as Multi-tenancy. This concept involves many users, for example employees of a business, sharing the capacity of a single instance of software (Ali 2015). This approach saves a lot of money for businesses since less hardware for processing is required, however it can pose some security problems. Since all activity is channelled through a single machine this can lead to overlaps in traffic. Within a single business this should not cause much of a concern since all employees have a mutual goal when using multi-tenancy, however should this traffic be intercepted by a malicious individual a security concern arises. For users of virtualisation software, establishing a secure connection to the software is paramount along with implementing the software correctly (Ali 2015).

4.4 Data Loss

With all the advantages Cloud computing provides such as accessibility and ease of use, the main associated disadvantage is the loss of data during writing, reading, storage, transmission, or processing operations. Data loss is often classified as accidental or malicious since attacks can also class as a loss of data. In the case of almost any business, the loss of data may mean a loss of revenue, as modern businesses thrive off of the use of digital storage. Current ways businesses counteract this drawback is using proper procedure when dealing with data. These are often found within the businesses policies. These can include anything from ejecting a memory stick before it is removed from a device to protect against corruption, to not leaving the office with any company data in case of accidental loss. Data corruption occurs when segments of data necessary to open a file are not found or have changed causing files such as images to display as distorted or disorderly (Lowe 2017). Error Correcting Codes (ECC) are often used to attempt to reverse the corruption of files. ECC memory is storage
used to carry out this process as it can detect internal corrupted data and fix it as it is processed (Fischer 2017). In order to prevent data loss via corruption, recommendations towards a data loss prevention plan will be drafted. The contents of the plan will include optimal storage and data transfer methods to coincide with other principles to assure data isn’t lost through other means.

After gathering, analysing and comparing the existing literature in this field, a series of recommendations will be made regarding the numerous areas of security within the Hybrid Cloud. To conclude the finding of the literature research, a summary containing overviews of the different areas of security will be provided. This section will recapitulate the recommendations made along side the evidence to support the choices made.

4.5 Legality and policies

Terms and Conditions (Ts&Cs) between the CSP and consumer must be considered when establishing an agreement between the two parties. It is vital to ascertain who is responsible for different attributes of the system. For example, who is responsible for ensuring that network security is operational and working effectively. If these areas are not covered, this can leave a hole in security that could be filled by clear communication. All CSPs should enforce a Service Level Agreement (SLA) before a consumer can begin the use of the company’s services, often in the form of a Ts&Cs click-to-accept form online (GOV 2017). This not only ensures that security and ethical procedure is maintained but also protects both parties from liability should anything happen to data when in the Cloud (Khan, K.M. 2010). Company policies also play a large role in the security of business operations. Employee use policies, when using a model such as Hybrid Cloud are the first line of defence. A good representation of the use of policies as a security defence is found in the “Defence in Depth” (Schneier 2006) shown in figure 2 of the appendix. Educating and informing employees on how to keep their activities on the cloud secure. Non-disclosure agreements are often enforced on employees by companies to restrict the flow of classified information from leaving the confines of the company. In this
particular case the employee agreement will also contain details of how to operate cloud services securely.
Literature Review
5.0 Literature review

The research into hybrid cloud computing already has a vast range of articles and papers dedicated to the area. With an even broader range of academic material dedicated to general cloud computing and the security it entails. The National Institute of Standards and Technologies (NIST) is the organization responsible for developing the standards and guidelines to uphold the security of the cloud for its operations. Mell gave a baseline definition of the hybrid cloud computing model and the services it provides, in an official paper released on behalf of NIST. They explain that the hybrid cloud model is a conformation of 2 or more types of cloud infrastructure. They go on to state that the 3 main infrastructures used in the hybrid cloud model are the public, private and community models (Mell et al. 2011).

5.1 Network and Application attacks

A Botnet is defined as the infection of a network of private computers with malicious software to send spam or shutdown a private network. These networks are easier to attack since they are semi private networks and yield more success since the private cloud network links directly with a public network (Daintith and Wright 2008). Sniffer attacks are also something mentioned by (Mathkunt 2014), this coincides with Kuyoro’s work as they both specify that server network attacks are a big risk to cloud security.

Distributed Denial of Service (DDoS) attacks are one of the most frequent attacks carried out on server-based services. The aim of a DDoS attack is to disable an infrastructural server network by sending thousands of data packets known as requests to a particular server (Daintith and Wright 2008). As put forward by Saied, a successful attack attempts to produce a network of botnets or zombies. These zombies are controlled by either a hacker or self-installed Trojans for example Flood-IM. To identify attacks Artificial Neural Network (ANN) was experimented with and returned a detection accuracy of 98%. ANN is a pattern recognition method used in high-dimensional parameter space. Fundamentally,
the detection method involves cross-referencing genuine packets released from real traffic with forged packets provided by zombies built in libraries (Saied et al. 2016). A different approach to deal with the same problem using path fingerprints has also been researched. This approach is addressed by embedding unique path fingerprints in each IP packet to trace the route the packet has taken and track its source (Lee and Shieh 2005). An additional paper, evaluating the use of ANN was found to agree with the finding of Lee and Shieh. The network was used to monitor a controlled test environment, which was compared with other monitoring networks without neural networking attributes. It was concluded that the use of ANN resulted in a 20% increase in monitoring capabilities and reliability (Grzonka et al. 2015).

In a blog written by G. Peterson named “Thinking persons’s guide to the Cloud, Part 3b”. It is mentioned that a number to methods used to prevent access control and ensure integrity and authenticity is kept intact. This is then expanded upon by Khan, when he states some businesses propose to use a mixture of claim-base access control, a security token system to be used with company hardware and federated identity approaches as security implementation (Khan, K.M., 2010). Risks such as data location, which will largely effect how the data is treated since different countries has different laws and policies as to how data should be stored and used (Kaur 2015). Kaur covers these immediate threats well by suggesting that cloud computing as a great asset to control, however loss over the data is a massive downside to cloud computing. Relating the locational security, the physical security of the premises on which the data servers are stored is equally as important as cyber security. It is recommended that data centres are secured against security breaches and other means of physical attacks (Mathkunt 2014).

The use of MetaCloudDataStorage security architecture is used to secure data dependent on sensitivity. Amazon Web Service (AWS) Cloud Trail is used in this proposed framework to process log files. It is pointed out that this security architecture can be used diversely since AWS Cloud Trail can be integrated with many APIs. Data is divided between different data centres or storage locations
depending on its sensitivity, deemed by the consumer. Data logging can also be recorded allowing the user to view who logged into access the data and when. The framework encrypts the file path of the big data, which is known as cryptographic virtual mapping. The path is encrypted instead of the big data itself since encrypting big data requires a lot of resources and time. Multiple copies of data will be stored in some location to ensure data integrity in case some data corrupts (Manogaran et al. 2016). This approach could be combined with different encryption algorithms such as the encryption method put forward by Chang et al. 2016 as both papers aim to provide safer techniques for storing and protecting data from 3rd parties.

5.1.1 Network and Application attacks Analysis

When providing security recommendations for network attacks, countermeasures must be implemented quickly and unambiguously. Efficient anti DDoS software must be implemented to prevent malicious activity from attacks. Access to permissions and especially sensitive data must be tightly restricted from unsolicited access. Access control such as a claim-based access used along with a token system would be a good example of a sound access control (Kaur 2015). All of the literature reviews recommending defensive techniques offer different approaches to prevent and filter DDoS connections. While the ANN approach looks to filter connections based on their data contents, using fingerprint paths, looks for the origin of the packet instead. Grzonka also found the use of ANN to increase reliability during detection. While the system used in this paper does differ from the use of ANN for Hybrid Cloud security, increased detection efficiency was still found. Different approaches have varying advantages and drawbacks making it important to implement the strongest system to support a hybrid cloud model.

5.2 Encryption

Kudtarkar pointed out that data security is much harder to maintain when implementing a hybrid cloud model due to having multiple CSPs. If there were to
be a security breach it would be near impossible to ascertain where the data breach occurred. It then goes on to recommend a security alternative using encryption and splitting to safeguard against breaches in data. Searchable Symmetric Encryption (SSE) is also proposed on the client side meaning the data would not be decrypted at the CPS, which would consequently lead to fewer breaches in security. This is described as a way to search through encrypted data for key words or text. Time would be saved since there wouldn’t be a need to pull an entire encrypted dataset to find the data required (Kudtarkar, P.P et al. 2015). Iqbal et al. 2016 supports this encryption approach as he concluded that endpoint encryption is a typical approach for users to protect their data. However, he does add that a change in method by hackers is beginning to focus more on application layer opposed to the network layer meaning endpoint encryption with be ineffective towards these kind of attacks. Another similar system for cloud data storage known as Zero-Knowledge System was put forward by Chang et al. 2016 which relates to a method of storage where the system knows nothing about the data it is holding such as contents of cloud files, metadata of user storage spaces, encryption keys, including passwords that protect cloud files and user storage spaces. This is just one of the 3 layers of security Chang et al. put forward to fully protect a system however, this would imply that used alone, it may be deemed less effective without the other 2 layers employed.

NIST has initiated the 256-bit Advance Encryption Standards also known as (AES) and began wider implementation around 1997. AES uses 3 different block ciphers AES-128, AES-192 and AES-256. Data is then encrypted and decrypted 1 block or 128 bits at a time using cryptographic keys of 128, 192 and 256-bits, respectively. Symmetric or secret-key ciphers use identical keys for encrypting and decrypting, so both the sender and the receiver must match the same secret key (Rouse, 2014). Iqbal agrees that this is a strong encryption method for securing outsourced data across a network (Iqbal et al. 2016).

Whilst most methods of data encryption are based around either symmetric or asymmetric algorithms, Rizvi et al. offers a fairly simple architecture using both symmetric and asymmetric algorithms. The proposed scheme uses a Trusted
Third Party (TTP) who is responsible for holding and communicating the secret key between parties. The Cloud Service User (CSU) encrypts the data with a secret key encryption algorithm before sending it to the cloud. The CSU then securely passes the secret key to the TTP where they will securely hold the key until decryption is necessary. The TTP is able to integrate the identity of both parties using the public key. When the CSP needs access to the data to perform data computations or check integrity, the secret key can be requested from the TTP. The identity of the CSP will be confirmed by the use of the public key. Whist being contradictory to the works of Chang et al. 2016 and Iqbal et al. 2016, this encryption approach still provides a secure method to storing data on the Cloud (Rizvi et al. 2014).

5.2.1 Encryption Analysis

In analysis of the Zero-Knowledge System put forward by Chang et al. 2016. In an ideal world this method of storage would be exceptionally secure and safe to use however it would be challenging to find a CSP that would be willing to store unknown data publically without having any benefit to storing the data. Many CSPs require the access of the data should they require it for their own security reasons. That said, this encryption approach would be very secure and a good recommendation if a CSP would agree to the terms. Rizvi et al. 2014 has argued an approach, which assures security by including an additional party. This would likely work more favourably for the CSP since it would mean they would have some insurance from the TTP that the data wouldn’t be malicious. However, the lack of access to the data may still negatively affect both the user and the CSP. For example, the process of integrity checking would be a lot more complex.

5.3 Virtualisation & Multi-Tenancy

Existing hybrid cloud challenges including multi-tenancy are discussed in a number of papers. Many authors define multi-tenancy similarly, as a single instance of software running on a server being used by a number of unique users
on their separate machines. This poses a data visibility threat to user operations. (Ali 2015). Another downside to multi-tenancy is known as “reputation fate sharing”. When a law abiding cloud tenant uses the machine and services as another tenant who doesn’t follow the same rules. This will negatively affect both of the users (Dillon 2010). Isolation of the tenants is very important when hosting a multi-tenancy system and should be considered over APIs and running services. This is because any attack on a virtual machine should not disturb other virtual machines on the operating system or server hosting the tenants (Iqbal et al. 2016). This is also referred to in Mathkunt’s paper as “virtualization security” but is based around the same idea that having a virtual instance running on a physical machine can cause security risks as attackers may gain control over the virtual environment (Mathkunt, 2014). He implies that this isn’t as much of a security issue and more of an infrastructural problem, however can be applied to security threats. This can also be contrasted with another idea, which suggests that hackers can use multi-tenancy to organise botnet scams (Kuyoro 2011). However, Che’s paper studying the security model of general cloud computing begins by agreeing that virtualisation within multi-tenancy is a very useful feature to control within the cloud computing model, but addresses it as more of a security risk. She suggests that it is important for companies to address the virtualisation technology that is implemented and apply the most secure depending on the product provided (Che et al. 2011). Furthermore, Mishra backs up these statements with statistics that suggest around 60% of virtualized servers will be less secure than the physical servers they replace. He then goes on to say that this number will have likely dropped to 30% by 2015 (Mishra, A. et al. 2013). Even if these statistics show a 30% drop in threat over the last 3 years, this would surely project that there is still a certain level of threat to the security of these virtual servers to this day. Whilst these papers don’t directly disagree with each other, they view the security risk of multi-tenancy virtualisation with differing degrees of threat to cloud security.

In order to reduce the occurrence of data loss through virtualisation, Virtual Machine (VM) replication has been looked into. Recovery Point Objective (RPO) is defined as the state of lost data after a disaster (Couto et al. 2015). Couto proposes a model, which produces a zero RPO when data loss occurs. This is
fundamentally achieved by monitoring VM state (e.g., disk, memory, CPU registers) at a given instant. Constant checkpoints are made when drastic changes occur with the state. If data is lost through system error, the latest checkpoint will be loaded recovering lost data (Couto et al. 2015).

Ren suggests that High Powered Computing (HPC) could be moved to a cloud infrastructure to save money while avoiding the drawbacks on performance. This is achieved by the lightweight virtualisation method that derives from the existing approach nOSV. Some statistics are then produced stating the average CPU utilization is only 3%~ 17% in Amazon EC2. The average CPU utilization in Google data centre is about 25%~ 35% and the average memory utilization is below 40%. The approach works by not virtualising all hardware resources but only the essential resources such as ACPI for CPU, E820 map for DRAM (Ren et al. 2017).

5.3.1 Virtualisation & Multi-Tenancy Analysis

Multi-tenancy should not leave any exploitable holes in security through the VMs used. When implementing security procedures for businesses, it is important that the users cloud software is fully secure from malicious users. The research above also shows that the infrastructural service also requires attention in the form of anti-virus or other fortification software to be completely secured for businesses. Since hackers can use multi-tenancy to organise botnet scams (Kuyoro 2011) security software would also mean that risk of malicious activity would be reduced. As long as all security measures are maintained, security should be sound, however as said by Mishra VMs aren’t generically as secure as the real hardware itself (Mishra, A. et al. 2013).

5.4 Data Loss

Zafar presented taxonomy of existing data schemes as well as pointing out weaknesses of the schemes as well as providing some insight into how these weaknesses can be improved. He identified data lock-in as a negative effect of storing data with a single CSP and then changing providers in the future. Since
CSPs use varying storage formats and store data in different drive sectors, moving data from one CSP to another can be a cause of data loss or corruption. Bad sectors on the Cloud drive may also lead to corrupted data if the CSP is unaware of any drive malfunctions. To resolve problems such as these, a data integrity scheme is recommended. A Deterministic approach can be taken which involves accessing 100% of a file to check for its integrity. This method is more accurate but doesn’t give a full field of view for other data. This method isn’t suitable for large files such as large archives since it takes too long to scan everything and isn’t affective. For larger sets of data, a probabilistic scheme is used which scans random blocks of data for integrity issues. This isn’t as accurate but when used on a large data set is more efficient (Zafar et al. 2017). A similar data-auditing scheme is put forward in another paper, similar to Zafar’s, which successfully analyses data copies within Cloud storage. This proposal analyses copies of files instead of the file itself so that the origin data is not disturbed. The system uses a trusted third party to verify the data in question and to also insure that data is not manipulated by either party. This approach uses Fully Homomorphic Encryption, which allows the cipher text to be analysed without compromising its content (Yi et al. 2017).

Whilst running these schemes is effective to ensure data loss and corruption is prevented, these schemes can also fall victim to malicious attacks. Bad CSPs can carry out a tag forgery attack, which involves producing faked result tags and makes damaged data seem intact. This attack is usually carried out to prevent auditing and deceive the verifiers (Zafar et al. 2017). This data scheme also provides some useful tools for editing and deleting data in blocks without having to download the entire file from the Cloud. These processes are done using similar algorithms to checking the integrity of the data. For editing the necessary blocks are pulled out and the contents of the specific block is appended, or deleted using the deletion algorithm.

Yu proposes an identity-based cloud data integrity checking protocol (ID-CDIC) from an RSA signature. This checking system is said to be more efficient than other data integrity checking systems since it avoids certificate management,
directly making the system more simple (Yu et al. 2016). This system would link more specifically to Zafar's work, offering a specific example of a data integrity scheme and its unambiguous components. It goes as far as to even offer a prototype of a 1D-CDIC security model. The mention of Amazon’s EC2 cloud server crash of 2011 is made to justify the importance of data integrity systems. In the literature review of this paper, a mention is made to Ateniese who proposed a notion of Provable Data Possession (PDP), a system that authenticates the true owner of outsourced data (Ateniese et al. 2007).

5.4.1 Loss of Data Analysis

While all authors put forward a slightly different approach to avoid data loss, increasing integrity is always the outcome. We can interpret that data should be moved and accessed only when required and while reading methods vary, data integrity and availability must be top priorities. Integrity checking systems seem to be popular in the process of data loss prevention, which should be taken into account when giving recommendations.

5.5 Legality and Policies

As stated by Brodkin, when it comes to the public storing or processing their data through the cloud, it is not always the provider who is responsible for the security and integrity of the their data. Regulatory compliance is an important legal aspect of cloud computing and poses a vast risk if users are unfamiliar with its terms. In order to overcome these security risks, a CIA (confidentiality, integrity and availability) model can be used to break down the threats (Brodkin 2008). Kaufman backs up the usage of this model to address regulatory compliance. Kaufman then goes on to suggest that in order to meet the requirements of the CIA model, the provider must offer the minimum capabilities of: A tested encryption scheme to ensure that the shared storage environment safeguards all data, stringent access controls to prevent unauthorized access to the data and finally, scheduled data backup and safe storage of the backup media (Kaufman 2009). In another area of legality behind Cloud Computing, service
level agreements (SLAs) are often implemented by providers in order to establish trust with their users while also outlining security measures that must be met by both parties (Khan, K.M. 2010).

A 2012 study on Cloud Computing and its effects on legal structure, pointed out the 5 essential elements of accountability. These include: Organisation commitment to accountability and adoption of internal policies consistent with external criteria; Mechanisms to put privacy policies into effect, including tools, training and education; Systems for internal, on going oversight and assurance reviews and external verification; Transparency and mechanisms for individual participation; and Means for remediation and external enforcement (Cheng and Lai 2012). This mentions that Cloud Computing privacy laws fall under a grey area due to the diversity of the services they offer. Cloud Storage would have completely different compliances compared to a communication service hosted on the cloud, which have to comply with the “The Stored Communications Act” in the US. There is also speculation that it is unclear whether the CSP is a data collector or a data processor since there are different regulations and it’s also difficult to regulate since data collection laws differ in the US compared to Europe. This also causes issues if the CSP is labelled as both since regulations contradict each other (Cheng and Lai 2012). This concludes that the impacts made by Cloud Computing create completely new legal issues, or at least make pre-existing legal issues prominent. And legal bodies have their work cut out for them ensuring an airtight legal system is in place for CC and its security concerns.

Data protection laws protect data owners from exposure to a certain extent. CSPs must abide by these laws at the risk of prosecution. Gray discusses the conflicts involved relating data protection laws to Cloud Computing. Cloud users may argue a contract breech has been made if data that belongs to them has been given to other parties e.g. advertising firms. He goes on to say the legal course of action would be to examine precise clauses in the Cloud terms of use created by the CSP. It is mentioned that legal action is executed in different ways depending on the jurisdiction. In Europe, the relevant provisions are found in Regulation (EC) No598/2008 of the European Parliament and of the Council on 17 June 2009. If the user indicates that the breech of privacy amounts to a tort, meaning an
infringement of a right, the case becomes more complex. He states in the UK, Cloud data breeches aren’t traditionally recognised as a tort of privacy. UK legislation takes into account the laws of where the data breech occurs, if it occurs in multiples places, it considers the laws of the location where the most significant breech occurred. He concludes with an overview and opinions on the laws in the EU. The EU allows flexibility to each of its members to maintain their own data protection laws. However does mention that legislative privacy protection can be worked around or undermined via loopholes (Gray 2013).

In a paper discussing the industrial and economic factors effecting the development of cloud computing in China. It is remarked that a lack of strong privacy and data protection law hinders the full potential of China’s industry. He also suggests that CPS oligopoly such as Microsoft and Oracle lobby for a streamlining of the European Union’s fragmented national data protection laws. This is due to China’s weak civil state which prevents them expanding their services to the US and Europe (Kshetri 2016).

In a paper by Carlin, S. et al., 2011, it is depicted that a large security risk lies in the fact that the demand for cloud computing will overwhelm the network providers causing a lapse in security due to enormous amounts of network traffic.

5.5.1 Legality and Policies Analysis

When Kshetri remarked about China not being able to reach their full potential, he doesn’t consider the security consequences of what would happen should Chinese industry not fully come up to speed with the western world’s security and privacy policies. Integration of Chinese industry could cause a large unintentional breech in security. I disagree with Carlin since many large network providers have contingencies and necessary funding in place for when expansion is required. The discussion of data protection laws by Gray mentions European laws however doesn’t consider the recent effects of Britain leaving the EU. This may effect the data protection laws previously implemented.
Evaluation & Recommendations
6.0 Evaluation & Recommendations

From the literature reviewed, we can see that in solution to the security threats posed by the Hybrid Cloud, there are several methods of solving problems in their specific areas. These areas will be evaluated individually and summarised at the end of each section with the best possible solution. It is problematic to impose rigid security using a hybrid model since we are dealing with an overlap of models, both private and public (Kudtarkar, P.P et al. 2015). This would indicate that security decisions made, need to be airtight and appropriately applied.

6.1 Network and Application attacks

This section will cover both network and application level attacks and how they can be prevented. When addressing network vulnerabilities, DDoS, one of the most likely attacks in this case has been considered so a more in-depth recommendation can be made (Microsoft 2015). When addressing application attacks, the majority of research suggested access restriction approaches. For this reason, access control has been exclusively evaluated to narrow the broad field of study.

6.1.1 DDoS attacks

Servers hosting Cloud infrastructure, company platforms and software are huge assets for the majority of modern businesses. Many of these networks store valuable assets including personal or financial data. While many view this as a secure data vault, not to be accessed by any unauthorised personnel, others see it as an opportunity to gain valuable information. On a different side of things, some malicious individuals or groups aim to take down networks and hosted services to hinder the operations of a business. The most common approach to achieve this is through DDoS attacks. Once many an attacker has infected enough unique machines to distribute the attack, packets of data are sent to the targeted
server in order to flood it with connections. To prevent this type of attack we must first identify the botnet connections from genuine connections to the network (McDowell 2009).

Research has been conducted using neural networks and self-learning machines to identify the difference between zombie connections and genuine connection (Saied et al. 2016). This approach involves providing an Artificial Neural Network (ANN) with the 2 different types of connection, zombie and genuine. Hidden layers are then applied to the computational learning process before thousands of varying connections are provided to the network so the system can begin to differentiate between the 2 types. The main differences between the types of connections are the packets received by the server. When packets are received from a zombie connection the DDoS tool randomly generates their packet data such as IP data. This is what the intelligent system must decipher between. While the paper researched both Transmission Control Protocol (TCP) and User Datagram Protocol (UDP), for this case we will only be looking at the TCP layer. The research finds that 2 hidden layers are optimal to the network detecting the correct outcome. Any more or less and the system is less accurate (Grzonka et al. 2015).

*Fig. 3*

Figure 3 represents the optimal neural network used to detect the DDoS packets across the TCP layer (Saied et al. 2016).
This approach is a valuable asset to many companies since the network can constantly be updated and continue its learning process as it operates. The system will require few manual algorithm updates, as it will be able to evolve as botnets become harder to differentiate. This concept can also be applied across many layers from network layers to application layers.

Disadvantages to this system however include the extensive learning time required to distinguish between genuine and artificial connections. There is also the chance that some zombie connections could be misinterpreted as genuine connections. If a company is looking to roll out a system to prevent Cloud DDoS attacks with instant efficiency, this system may not be suitable. Implementation of a neural network also takes trial and error with numerous iteration periods extending the time taken for the system to work effectively (Saied et al. 2016).

This approach compared to other DDoS countermeasures is more highly recommended due to the future capabilities it holds. Compared with other analysed research such as fingerprint path detection (Lee and Shieh 2005), it offers companies a more diverse solution to fight against attacks. Whereas most generic DDoS firewalls search for connection under pre-set parameters, the ANN redefines its own parameters as it sorts between the 2 types of connections. It would appear ANN is also a more head-on approach to fighting the attack, IP paths may not always be traced if they are well hidden also making ANN not only more direct but more accurate over a long period. This has been proven in research compared with non-ANN methods and yields a higher detection rate making this the superior DDoS detection software (Grzonka et al. 2015).

6.1.2 Data access control

If a malicious individual were to bypass primary defences and gain access to the network, assuring they do not gain access to data within the Cloud is the next stage. Access control defines who has permission to areas of storage and who doesn’t.
The proposed concept to suit a Hybrid Cloud Model is MetaCloudDataStorage. This idea stores data dependant on its sensitivity to the business. This is well integrated into a Hybrid IaaS model since more secure data including personal and financial information can be stored in private sectors of the cloud with higher levels of security, whereas less sensitive data is stored in more public sectors provided by public CSPs. Data sensitivity is ranked Sensitive, Critical or Normal. This system allows tiers of clearance to be employed when accessing data, which not only keep sensitive data locked down but also helps with the data flow within a company. Restricting access to data is important within the Cloud but also allowing data to flow is vital for business to thrive. MetaCloudDataStorage offers the perfect balance for this. Data can only be accessed in sensitive sectors if the user possesses the right credentials on the system. Their access ID is compared with an employee access database. Logging access to data is also recommended, as the identity of who has been accessing sectors of storage is a big asset when monitoring security. Backups of sensitive files are also made to ensure integrity.

As well as MetaCloudDataStorage, physical access control is also recommended to provide extra security for data. A security token system is proposed involving hardware tags used by employees to access data. These can be connected to company computers allowing access to sensitive data stored on the Cloud. The token stores a digital signature that is cross-referenced with a database of signatures allowing or denying access to the owner. An advantage of employing this mechanism is that restricted data cannot be accessed remotely from outside the organisation without a token, meaning hackers will have to gain possession of the token or the encrypted database before they can access any data (Manogaran et al. 2016).

All research carried out provided either software or a hardware access control. However, using both software and physical hardware security approaches fused together provides a more well rounded security solution. Once the user has entered their security token to access restricted data, they will have to enter their credentials on the system.
6.2 Encryption

Encrypting data is the backbone of privacy when it comes to protecting data. Cryptography is one of the oldest methods of hiding information and in this modern day it has developed extensively (Menezes 2002). Businesses should be encrypting the majority of their data dependant on its sensitivity. We will be evaluating several encryption methods before recommendations are made to best suit the Hybrid Cloud Environment. Selecting the process of how data is hashed is relative to our scenario. Encryption algorithms have been cracked and updated at a constant rate throughout the last 50 years so it is essential that a strong and lasting algorithm is chosen. The main points to consider when deciding this are the strength of the encryption method and the usability on a day-to-day basis. Encrypting and decrypting data to the cloud shouldn't be a time consuming task for companies to undergo.

Symmetric encryption methods yield good security with fast and easy processing times due to fewer stages of encryption and decryption. One of the most popular symmetric encryption algorithms includes AES, more specifically known as Rijndael AES, which is currently employed within the US government to protect sensitive data. AES uses block cipher encryption, which takes blocks of 128bits, 192bits and 256bits (National Institute of Standards and Technology 2001). Some other symmetric algorithms use stream ciphering however AES has been proven harder to reverse engineer due to its block cipher algorithm (Pelzl and Paar 2010). If we compare a well established algorithm such as AES to a less commonly known encryption system such as Twofish, which uses a Feistel network to encrypt data, we see that they both reach the same encryption outcome (Schneier 2004). Twofish however has a more complex algorithm, making it inferior to AES since the algorithm is more cumbersome, however achieves the same outcome.

Benefits of asymmetric encryption include the extra security of having 2 separate private keys and therefore not having to send and receive a shared key. Not only does this save the task of securely sending the key between 2 parties, it also means that every person has their unique key reducing the chance of leaking
keys. The 3 sets that support asymmetric algorithms are RSA, DSA, and Elliptic Curve. Due to algorithm simplicity similar to symmetric encryption, RSA is the most commonly use method. This method uses two large prime numbers to create cipher text near impossible to reverse engineer without the origin numbers (Rivest et al. 1978). Many differentiations in algorithms yield the same result with an alternate method of getting there. This is why its important to chose the most direct and least complex algorithm so that the process is as straightforward as possible.

SSE was also researched which uses a shared secret key, however when data is encrypted specific placeholders are used to easily identify the data when decryption is necessary. This means, entire datasets don't have to be pulled from storage to decrypt making other stored data more secure and speeding up the decryption process (Kudtarkar et al. 2015). Adding placeholders to encrypted data is likely to increase the size of the stored file giving this approach a potential drawback however. This does however require encryption to occur client side, which many CSPs do not agree with, as they may wish to access the data they are storing. Therefore, finding a CSP happy to abide by this method may be challenging.

While asymmetric and symmetric encryption have their advantages and drawbacks. Research has come to light to show that a hybrid of the 2 methods is more effective in securing data (Rizvi et al. 2014). This is accomplished by using public keys to encrypt data as well as the generation of a symmetric key to be passed to the 2nd party. For this method, the process involves using the symmetric key to encrypt the data and then, using the asymmetric public key, encrypt the symmetric key again to provide 2 layers of encryption to data. This encryption system also uses a Trusted Third Party (TTP) to pass the keys between the sender and receiver. This is a necessity between the CSP and company to avoid data manipulation occurring before or during transit. If a business hosts their own network a TTP will not be necessary as the keys are sent and delivered internally.
Comparing the use of this method to other more generic methods, it seems that hybrid encryption is suitable across both public and private platforms of the Cloud since they both run on similar infrastructures (Plummer et al. 2009). This suggests that there would be no compatibility issues integrating this encryption method into a hybrid Cloud model. Decisions to steer clear of symmetric or asymmetric stand-alone encryption lies in the fact that it would be less secure in this scenario, as businesses are aiming to provide as much security as possible to their sensitive data. Searchable Symmetric cryptosystems did provide similar amounts of security and complexity needed for our scenario, however was less suited as the complex algorithm was less efficient and would take more time to encrypt and decrypt data. The length of the algorithms when compared can prove this. Since optimality must also be considered when implementing encryption, a hybrid cryptosystem is more fitting for this purpose. Furthermore, employing this hybrid approach involves recommending both a symmetric and asymmetric method of encryption to be used together. Based on the research carried out, using a combination of AES symmetric encryption and RSA asymmetric encryption will give the most optimal form of encryption.
Figure 4 provides a graphical representation of how hybrid encryption is carried out. Cipher text is originally generated when the private key is applied to the file. Public key encryption is then used on the encrypted file to provide 2 layers of encryption. The receiver must ensure that public decryption occurs before the shared private key in order to access the original file.

### 6.3 Virtualisation & Multi-Tenancy

Running software within a virtual environment yields many benefits for companies adopting a Hybrid Cloud model. A reduced cost in hardware, maintenance and power usage mean money can be invested in other important aspects of business. A decrease in Total Cost of Ownership (TCO) is also a striking advantage. TCO is defined as the cost of installing new products or systems, in this case computing hardware and software (Han 2011). Delving further into the value of virtualisation is the increasing use of software multi-tenancy. This concept further reduces business overhead as less hardware is required to operate.
Figure 5 shows the relationship between hardware and Virtual environments used in virtualisation. As we can see, hardware resources are split between multiple VMs (Firoozjaei et al. 2017).

Unlike dividing usability between hardware, virtualisation divides the processing power of a larger machine to provide virtual environments for users. However, multiple machines are unusable when updates are required to hardware since multiple instances run off a single machine. Unstable software updates can be more devastating as many virtual environments are infected by potential faulty or insecure software (Firoozjaei et al. 2017). These vulnerabilities leave businesses open to application level attacks or even backdoor network access. Damage to the system can also be caused by malicious insiders. Having access to administrative permissions can lead to damages. External threats should be ruled out when dealing with Hybrid Cloud virtualisation. If incorrect network configurations are made, malicious outsiders must be considered a risk (Firoozjaei et al. 2017).

We can implement the same access control used for MetaCloudDataStorage in this scenario. Physical controls such as data tokens can be used as well as login
authentication to restrict access only to authorised users (Manogaran et al. 2016). Activity logging software installation on virtualisation hardware is likely to be the more direct approach.

VM backups also need to be considered when establishing secure virtualisation. Research would suggest that VM replication is an effective way to prevent data loss and in turn tighten security within virtualisation. Constant checkpoints are made by software when working within a VM, hardware statistics are monitored and if figures such as CPU usage peak then recovery checkpoints are made. If any hardware or software crashes occur, when systems are recovered, the latest checkpoint will be acquired reducing the amount of lost data (Couto et al. 2015). Backups are stored locally so are easily accessible when restoration is necessary.

When rolling out software updates and carrying out maintenance, assuring stability within the update is vital. Software needs to be tested before it can be made live to prevent any vulnerabilities being missed. Software should be protected by hosting the SaaS within a private cloud, restricting access from public connections. Anti DDoS software must be installed with any hardware, hosting virtual environments. This will provide front line protection against potential network attacks and reduce the regularity of software layer based attacks. Protection against insider attacks comes down to the use of role based access control. This grants layers of access to only essential employees within a business and mitigates insider threats. Monitoring of consumer activity whilst virtualisation software is in use is highly recommended. A service known as SecondSite is recommended when managing recovery checkpoints. This is used as an example in a paper by Couto and is required when setting recovery checkpoints (Couto et al. 2015).

6.4 Data Loss

Procuring a system to increase the integrity of data stored on the cloud is an important part of security within an organisation. A popular cloud backup and recovery company released a survey carried out in 2013 showed that 32% of
businesses lost data when using the cloud. 64% of data lost was due to accidental overwriting or accidental end user deletion (Thornton 2013). This is proof that data loss is a real issue when dealing with Cloud Storage. Companies often employ integrity schemes that access data to check for storage abnormalities and that files have not been altered. However, since there have been several methods to achieve this, and stored company data can differ in format and file size, choosing the most effective integrity scheme must be accomplished.

Research has been conducted into the field of probabilistic integrity checking which comprises of evaluating storage sectors chosen at random to find any bad sectors of data and recover the data if possible. Instead of searching each individual sector lineally, check blocks are placed in the data when it is stored and then accessed when the integrity of the data is checked, yielding a faster scanning time when many files are stored (Juels and Kaliski 2007). See Figure 6 of the appendix for a diagram explaining the use of probabilistic data checking. Another researched method of integrity checking, Deterministic checking has also been investigated. This protocol accesses every sector individually, however takes much longer for the integrity of data to be checked. This approach is more effective with bigger files large instances of files, as it is necessary for every element of the file to be accessible for it to work (Zafar et al. 2017).

Data lock-in is also identified by Zafar as a problem experienced when data is transferred from one CSP to another. If data integrity is not checked beforehand, files can become corrupt when the data is transferred between storage sectors. Also taken into consideration, was Identity-Based Cloud Data Integrity Checking (ID-CDIC). This protocol favours the use of unique identities for users of the Cloud, over traditional certificates used to identify data origins, making the process more efficient. Identities are given to each user so that when data is stored, its origin can be identified, certificates do not need to be generated every time data is stored reducing the resources needed (Yu et al. 2016).

Tying in with encryption, data integrity must be analysed, while maintaining confidentiality, Fully Homomorphic Encryption protects data contents while still providing data patterns to allow integrity checking to still occur. This is a good
approach if parties do not trust each other or a TTP cannot be established (Yi et al. 2017). This approach is less secure than hybrid encryption, however, as an alternate encryption method for organisations will less to spend on security, this can be recommended since data integrity checking will be made much easier to achieve (Stuntz 2010). However, using this approach on top of the recommended data encryption will overcomplicate the encryption system. If data is decrypted in the wrong order then this will lead to further corruption and data loss.

For organisations that have local private cloud servers located on their premises, backing up privately stored data is a well-known endeavour. However newly founded businesses or companies with less technological understanding may not understand the importance of backups. Since a 3rd party hosts Public Cloud services, the control of physical hardware lies with the CSP, which can be seen in figure 1 of the appendix. Private storage however, can be controlled from in-house by organisations. This extra control means data can be stored as desired, e.g. more valuable data stored on several drives. This is why it is recommended more valuable data is stored on Private opposed to Public Cloud.

It is highly recommended for all organisations to employ some sort of data protection scheme. For companies storing large amounts of smaller, independent files on the Cloud, such as records or archives, probabilistic integrity checking is recommended. This system will be faster and more efficient when certification on every uploaded file isn’t required. If data stored on the cloud consists of larger files such as program or software files, the use of Deterministic checking is advised. To avoid issues such as data lock-in, organizations should employ a single, reliable CSP for their cloud services. While ID-CDIC is the more direct method of integrity checking compared to systems using certification, it is the less generic approach. When insuring data hasn’t been lost, it is important that time is taken and therefore a faster method isn’t needed in this case.
6.5 Legality and Policies

The first level of enforced security within organisations often comes in the form of company policies. All employees adhere to contracts at the beginning of employment, providing employers opportunity to impose measures of what should and shouldn't be done within a business. Employees and customers coming into contact with Hybrid Cloud should be made aware of the security policies put in place.

When a service providing party uses the services of another party, parameters need to be set on the use of the services. This is often addressed using a Service Level Agreement (SLA). The use of this agreement addresses quality of service, availability and responsibilities between the 2 parties (Wieder, P et al. 2011). For organisations employing the use of Hybrid Cloud, this agreement will be between the CSP and Hybrid Cloud user. Often agreements within SLAs are preset by the service provider making negotiating on the service difficult for the 3rd party. For the negotiation of services an Operational Level Agreement (OLA) is used to determine the relationship between the 2 parties. Agreements to the responsibility of data must be specifically determined within this agreement so that data loss or breach of data doesn’t occur within regulatory compliance of the Data Protection Act 1998 (Brodkin 2008). Availability of service should be specifically arranged to cater for any problems that arise with the service, for example server downtime. The CSP should assure access to cloud storage and services are available at all times. As previously stated, grey areas do appear due to the diversity of services offered by CSPs so ascertaining the responsibilities of both parties is necessary to protect data.

Company employee policy usage is also a suitable method of conveying security throughout a company. When employees gain a role at an organisation that involves handing sensitive data, a Non-Disclosure Agreement (NDA) is a reliable way to control the flow of data in and out of a business. The fundamentals of this agreement outline what a party can and can't do with data stored by an organisation, along with specific guidelines of how data should be handled. This
legal agreement can lead to prosecution if not abided by, making it a solid line of defence when protecting data (Radack 1994). An example of clauses within a NDA would include physical transit of data, outside the premises of the association. Taking a memory stick or company device outside of the office would class as a breach of this clause. NDA clauses also cover who information held by an organisation can be discussed with. Often people outside of the NDA loop cannot be told creating a security barrier to the outside world.

An agreement with the end user and company providing the service should also be made in the form of a Terms and Conditions (T&C) agreement. Online or software services often provide this contract in the form of a Click-to-Accept form. Contents of this agreement often state that the provider of the service or product will likely store data provided by the consumer. Data protection laws customarily stated at this point allowing the user to see that their data will be handled properly in accordance with the Data Protection Act. Bring Your Own Device (BYOD) policies are being more common within businesses to promote productivity, however should be revised within organisations that handle sensitive data as they may pose a security risk.

Access control for individuals is an effective method of restricting who can handle data within an organisation. Different departments within organisations should only be dealing with information specific to them. Access control systems (ACM) are used within many companies that store sensitive data and can be used within computational systems and physical premises, for example to control who enters server rooms and other restricted areas. This can also come in the form of security within an organisation, whether it be CCTV or security guards.

Data security should be the most significant element of security when arranging an agreement with a CSP. A trusted CSP should be employed and a healthy relationship maintained to prevent problems. If the consumer of the Cloud is responsible for the integrity checking of data, the contractual agreement should show this to reduce uncertainty. This also goes for aspects such as the moving of physical drives. If the CSP moves any storage devices causing the corruption of
data, the party accountable must be established. Contractual agreements invoke communication between parties making problems such as data loss or breaches less likely, leading the higher security within the Cloud. Further security can be reached by implementation of a clear NDA between users of Hybrid Cloud data and the company. Recommended restrictions also include strict control of external access within companies such as social media access. Company terms and conditions should be updated as regularly as necessary to ensure that they meet the guidelines of revised law.

Organisations need to categorise the sensitivity of information they hold and set relative security standards. While higher security may hinder workflow, safekeeping of assets must be addressed first. For companies adopting the Hybrid Cloud approach, it is advised that the data protection policy be divided between public and private cloud networks. As previously mentioned, more sensitive data would be more suited to private cloud networks while more generic data be found on the public network. Having different policies when dealing with the different networks means that precautions don't need to be taken when they aren't necessary. Fundamentally, this compartmentalisation will save time and increase workflow while still providing security within necessary sectors. A live example of this could be that company data that isn't considered secure is permitted in external locations, however more delicate data cannot leave the network.

Access control policies are also very important when protecting data. It is advised that different employees are given specific levels of clearance so that data can remain in the correct hands. Reporting policies are also a reliable way to ensure users only access necessary data, this can be done via logging when data is accessed and by whom. Enforcing punishment for unauthorised access is recommended.
Conclusion
7.0 Conclusion

In this paper, 5 main areas of security are outlined and should be addressed by businesses when employing a Hybrid Cloud model. These 5 areas include Network and Application attacks, Encryption, Data Loss, Virtualisation and multi-tenancy, Legality and Policies. Research has been conducted into these fields and a review of the existing literature has been assembled. Following research, justified recommendations have been made addressing how to reduce security threats across the 5 fields. The 5 sections will be summarised below.

7.1 Network and Application attacks

As addressed in the previous section, attacks from both network and application layers cannot be tolerated. Mention of existing research into attack prevention was discussed and evaluated. Methods of preventing DDoS attacks were researched across a number of academic papers to determine the best course of action against these attacks. Various major components of DDoS attacks were brought to light using pre-existing academic material such as botnets and how they are used to take down networks (Daintith and Wright 2008).

After establishing the fundamental elements of these attacks, researching existing variations of countermeasures was the next step. Anti-DDoS methods such as fingerprint path detection and ANN were evaluated and it was recommended that ANN be integrated into businesses to detect network attacks. The reason this detection algorithm was chosen over others was due to it being, over time, more efficient for organisations to upkeep and would require less time and attention than other non-learning anti-DDoS methods.

Although the initial cost may be higher than other methods to reduce network attacks, over time it is determined that ANN will be the most effective and cost efficient for businesses. Application level attacks were also considered, focusing on the control of access to data through applications. MetaCloudDataStorage was investigated to protect software access. The decision was to use this approach
because it provided a sufficient level of security whilst still allowing day-to-day business to flow.

Physical security in the form of surveillance and a security token system was also recommended to provide a well-rounded security approach when protecting against software level attacks. To summarise, in this section we have provided means of defending against network and software layer attacks, as well as providing some additional physical security to protect company data and services hosted by Hybrid Cloud.

7.2 Encryption

An approach to secure encryption was also proposed within this report. Research consisted of discovering effective and efficient methods of encrypting data, which fell into 2 categories of symmetric and asymmetric encryption. Symmetric encryption research turned up popular encryption algorithms such as AES used by government bodies in the US such as the NSA. Twofish, another symmetric encryption method was explored to reveal that both approaches yielded the same encryption result, with AES achieving this result via a more simplified process. Asymmetric encryption researched showed that RSA was the most efficient public key algorithm.

Searchable Symmetric encryption was also considered for recommendation, however, due to the popular use of AES and RSA, it would be more viable to go with a simple, strong and more commonly used encryption method. Both encryption algorithms were recommended in the form of hybrid encryption to enforce extra data security. Hybrid encryption is recommended over a single choice of encryption to cater for sensitive data stored by organisations. Although this encryption is more costly than a non-hybrid encryption method, it better suits business needs for maximum security. Recommendations were also based on the fact that encryption and decryption should be as efficient and time effective as possible while still being secure. This is why a balance between secure hybrid encryption and 2 simplistic algorithms were chosen.

For organisations with cost limitations in regards to security, hybrid encryption may not be possible to obtain. In this case, a simpler and cheaper approach to
encryption, Fully Homomorphic Encryption may be adopted. While not as secure as a hybrid approach, security and data integrity can still be maintained.

7.3 Virtualisation & Multi-Tenancy

Virtualisation is a very useful asset to be employed by companies following the Hybrid Cloud model due to using less hardware. Researched security solutions include the use of VM recovery checkpoint setting to reduce the chance of data loss as a result of a hardware or software malfunction. Services such as SecondSite are useful when taking recovery checkpoints as it assists with storage management. All updates installed to Virtualisation software should be vigorously tested before installation to prevent any vulnerability affecting the Cloud services. Access control to virtualised software is recommended to prevent unauthorised access to data. Activity logging helps monitor that the correct employees are accessing data and that data is being handled in ways that comply with company policy. The access control company policy applies here to enforce security when virtualisation is used. Anti DDoS recommendations should also be applied to this area as the hardware running Virtualisation software will also need to be protected from software layer attacks, which means installation of any network attack software will be necessary. This will also assist any network related attacks since Virtualisation requires network support. This strictly relates to use on Hybrid Cloud models since many connections from public areas will be expected as well as private connections.

7.4 Data loss

Addressing the accidental loss of data is an important element of Hybrid Cloud security considering this affects almost every business within it’s lifetime (Thornton 2013). In this section, recommendations are made dependant on the type of data businesses are storing. Research was conducted into ID-CDIC as an approach to data checking compared to other methods such as probabilistic or deterministic. Probabilistic integrity checking is recommended for businesses when smaller data files are being checked. When larger files are being used,
Research points towards the use of Deterministic checking being recommended for larger data stored on Hybrid Cloud. It is also advised that both types of checking be implemented so that the most accurate approach of integrity checking can be applied. Fully Homomorphic Encryption is recommended for use with companies that do not employ hybrid encryption. This encryption method allows for integrity checking to occur even when data is encrypted making decryption redundant for this process. Categorizing data on size could be a useful approach when using both methods because it would fit the architecture of the Hybrid Cloud model. Ensuring that only a single, reliable CSP is employed can be suggested so that data lock-in does not occur. Some aspects of data loss can be linked into virtualization including the use of VM backups so that if there is a hardware or software error, a checkpoint may be rolled back to and data loss will be reduced.

7.5 Legality & Policies

Employees within a business must be considered a security threat when organisations require high levels of security. Recommendations concerning policies and contractual agreements with CSPs was considered to prevent any vulnerabilities. Initial research into Data Protection laws was considered so businesses know what services CSPs should be offering by law. Investigation also revealed the importance of SLAs and OLAs when working with an external party. It was recommended that responsibilities such as integrity checking and other administrative duties be determined within an agreement so no accidental loss of data occurs. Stakeholder non-disclose agreements were also highly recommended, as they provide a legal seal on any information people acquire which can’t be protected any other way. These agreements can be used to prosecute anyone who breaches the contract providing incentive for employees to not leak information.

Access control policies should be implemented which ties in closely with the recommendations made against application layer attacks. Using a combination of employee based policies such as restricting external access through company
computers and installed software to restrict access, restricting data breaches from employees can be much easier to achieve. Categorisation of data sensitivity is also necessary so that employees know what data they should or shouldn’t be accessing. This also links back in with data access and encryption, since more sensitive data can be more safely secured while less sensitive data can adopt a more relaxed approach to security. This will lessen the disruption of time taken accessing data that isn’t necessarily required. Categorisation also suggests sensitive data be stored on Private Cloud systems as this system can be more strictly controlled and therefore be less of a liability.

7.6 Summary

The 5 areas addressed were chosen since many security elements overlap making them easier to understand and integrate into the Hybrid model. All recommendations made have considered the Hybrid Cloud infrastructure that businesses would be using, this would suggest why addressing security risks has been taken highly into consideration. While the security measures of businesses will be dependant on the type of data they will be holding, the recommendations given accommodate for a variety of data stored by organisations, for example, large data or smaller data. Depending on the data, a variable amount of security is required. This paper aims to provide a secure solution to data and services hosted on the Hybrid Cloud while also maintaining efficiency and ease of use. Security should minimally hinder the performance and operation of a business or organisation, which has been considered where recommendations were made.
8.0 Future Research

As Cloud Computing is a forever changing concept, security is likely to evolve as time goes on. Methods that are suitable for today's cloud technology may be made redundant in the future. As attacks against Cloud systems become more advanced, protection will also need to be reengineered to counteract these attacks in the future. Servers will likely have to withstand more traffic, requiring more processing power and storage. If new ways to attack a network are developed, this will also need to be accounted for.

While only 5 areas of Hybrid Cloud security were discussed in this paper, there are many other areas that require securing. Making sure network ports are secure when connecting to a server is a crucial element of network security. Further research could be undertaken into other infectious entities that effect software and networks including worms and Trojans. Further recommendations could be made for anti-virus software to fight these infections.
9.0 Appendix

Fig 1.

A visualisation of service management from CSP to consumer.

Fig 2.

Defence in depth model with labelling each layer of security to protect data.
An example of how probabilistic data checking accesses data segments in an orderly way when evaluating the integrity of smaller files. In this diagram, y1, y2, y3, and y4 represent the data being checked against its storage certificates X1, X2 and X3. All b labels represent random order that data integrity is analysed.

When undertaking a research or enterprise project, Cardiff Met staff and students are obliged to complete this form in order that the ethics implications of that project may be considered. **If the project requires ethics approval from an external agency (e.g., NHS),** you will not need to seek additional ethics approval from Cardiff Met. You should however complete Part One of this form and attach a copy of your ethics letter(s) of approval in order that your School has a record of the project. The document **Ethics application guidance notes** will help you complete this form. It is available from the Cardiff Met website. The School or Unit in which you are based may also have produced some guidance documents, please consult your supervisor or School Ethics Coordinator. Once you have completed the form, sign the declaration and forward to the appropriate person(s) in your School or Unit.
PLEASE NOTE:
Participant recruitment or data collection MUST NOT commence until ethics approval has been obtained.

PART ONE

Ethics Number - 2016D0339

<table>
<thead>
<tr>
<th>Name of applicant:</th>
<th>Samuel Bean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervisor (if student project):</td>
<td>Paul Angel</td>
</tr>
<tr>
<td>School / Unit:</td>
<td>School of Management</td>
</tr>
<tr>
<td>Student number (if applicable):</td>
<td>20057331</td>
</tr>
<tr>
<td>Programme enrolled on (if applicable):</td>
<td>BSc (Hons) Software Engineering</td>
</tr>
<tr>
<td>Project Title:</td>
<td>Security Recommendations for organisations employing a Hybrid Cloud model</td>
</tr>
<tr>
<td>Expected start date of data collection:</td>
<td>N/A</td>
</tr>
<tr>
<td>Approximate duration of data collection:</td>
<td>N/A</td>
</tr>
<tr>
<td>Funding Body (if applicable):</td>
<td>N/A</td>
</tr>
<tr>
<td>Other researcher(s) working on the project:</td>
<td>N/A</td>
</tr>
<tr>
<td>Will the study involve NHS patients or staff?</td>
<td>No</td>
</tr>
<tr>
<td>Will the study involve human samples and/or human cell lines?</td>
<td>No</td>
</tr>
</tbody>
</table>

Does your project fall entirely within one of the following categories:

<table>
<thead>
<tr>
<th>Category</th>
<th>Yes/No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper based, involving only documents in the public domain</td>
<td>Yes</td>
</tr>
<tr>
<td>Laboratory based, not involving human participants or human samples</td>
<td>No</td>
</tr>
<tr>
<td>Practice based not involving human participants (eg curatorial, practice audit)</td>
<td>No</td>
</tr>
<tr>
<td>Compulsory projects in professional practice (eg Initial Teacher Education)</td>
<td>No</td>
</tr>
<tr>
<td>A project for which external approval has been obtained (e.g., NHS)</td>
<td>No</td>
</tr>
</tbody>
</table>

If you have answered YES to any of these questions, expand on your answer in the non-technical summary. No further information regarding your project is required.

If you have answered NO to all of these questions, you must complete Part 2 of this form

In no more than 150 words, give a non-technical summary of the project

The main security risks towards hybrid cloud computing will be named and discussed followed by recommendations concerning how to prevent these risks and deal with them once they have occurred. The focus of this research paper will be involve insider attacks, which will focus on the protection of hardware and physical devices such as passwords. Another area that will be addressed is the software side of hybrid cloud computing which includes making sure relevant firewalls are in place to negate network attacks. Finally we will look at the hybrid model and locate downfalls in the hybrid system for example the grey areas between public and private cloud storage and computing.

DECLARATION:
I confirm that this project conforms with the Cardiff Met Research Governance Framework

I confirm that I will abide by the Cardiff Met requirements regarding confidentiality and anonymity when conducting this project.

STUDENTS: I confirm that I will not disclose any information about this project without the prior approval of my supervisor.

Signature of the applicant: Date: 04/11/2016

FOR STUDENT PROJECTS ONLY

Name of supervisor: Paul Angel Date: 08/11/2016

Signature of supervisor:
### Research Ethics Committee use only

<table>
<thead>
<tr>
<th>Decision reached:</th>
<th>Project approved</th>
<th>Project approved in principle</th>
<th>Decision deferred</th>
<th>Project not approved</th>
<th>Project rejected</th>
</tr>
</thead>
</table>

**Project reference number:** Click here to enter text.

**Name:** Click here to enter text.  
**Date:** Click here to enter a date.

**Signature:**

**Details of any conditions upon which approval is dependant:** 
Click here to enter text.

### PART TWO

#### A RESEARCH DESIGN

**A1** Will you be using an approved protocol in your project?  
No

**A2** If yes, please state the name and code of the approved protocol to be used

N/A

**A3** Describe the research design to be used in your project

Necessary protocol will be used when referencing sources of information not belonging to myself. During secondary data collection all external sources will be referenced. All of the literature review will be referenced using the Harvard referencing format unless it is my own work.

**A4** Will the project involve deceptive or covert research?  
No

**A5** If yes, give a rationale for the use of deceptive or covert research

Click here to enter text.

**A6** Will the project have security sensitive implications?  
No

**A7** If yes, please explain what they are and the measures that are proposed to address them

Click here to enter text.

### B PREVIOUS EXPERIENCE

**B1** What previous experience of research involving human participants relevant to this project do you have?  
None

**B2** Student project only  
What previous experience of research involving human participants relevant to this project does your supervisor have?  
N/A

---

1 An Approved Protocol is one which has been approved by Cardiff Met to be used under supervision of designated members of staff; a list of approved protocols can be found on the Cardiff Met website here.
C POTENTIAL RISKS

C1 What potential risks do you foresee?
Not citing sources correctly may lead to confusion and potentially accidental plagiarism.

C2 How will you deal with the potential risks?
By assuring that all work that is not my own is referenced according to the referencing guidelines.

When submitting your application you MUST attach a copy of the following:
- All information sheets
- Consent/assent form(s)

An exemplar information sheet and participant consent form are available from the Research section of the Cardiff Met website.

DEVOLVED ETHICS APPROVAL APPLICATION SUMMARY

Student Name: Samuel Bean   Student Number: st20057331

Module Name:___________________   Module Number: __________

Programme Name: BSc (hons) Software Engineering   Supervisor Name: Paul Angel

<table>
<thead>
<tr>
<th>To be completed by student and supervisor before submission to Ethics Approval Panel</th>
<th>Student Signature;</th>
<th>Supervisor Signature;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application for ethics approval</td>
<td>[x]</td>
<td>[ ]</td>
</tr>
<tr>
<td>Participant information sheet</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>Participant consent form</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>Pilot interview/s</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>Pilot questionnaire/s</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>Letter/s to participating organisation/s</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>Confirmation of interviewee participation</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td></td>
<td>[x]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

First Submission [x]    Resubmission [ ]

Date: 26/11/2016

For use by the devolved ethics approval panel:

Panel Members Name Signature

Module leader, Chair:

Supervisor: ___________________________ ___________________________

CSM Ethics Committee Representative: ___________________________
<table>
<thead>
<tr>
<th>Outcome</th>
<th></th>
<th>Reference number issued:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Approved</td>
<td>[ ]</td>
<td></td>
</tr>
<tr>
<td>Chair’s Action</td>
<td>[ ]</td>
<td></td>
</tr>
<tr>
<td>Application not Approved</td>
<td>[ ]</td>
<td></td>
</tr>
</tbody>
</table>

Comments for projects not fully approved:

The original to be retained by the supervisor and a copy given to the student and module leader.
In the case of a resubmission being required this original form should be submitted with the resubmission not a new, blank, one.
10.0 References


• Lowe, E. 2005. ZFS saves the day(-ta)! (/var/crash/elowe) [Online]. Available at: https://blogs.oracle.com/elowe/entry/zfs_saves_the_day_ta [Accessed: 3 April 2017].


• Dillon, T. 2010 24th IEEE International Conference on Advanced Information Networking and Applications


• Mell, P. and Grance, T., 2011. The NIST definition of cloud computing.


423-434.


• National Institute of Standards and Technology, 2001. ADVANCED ENCRYPTION STANDARD (AES).


