WBS

AY 2019

BUSINESS DRIVERS OF INNOVATION IN ENCRYPTION

Use Cases for Homomorphic Encryption

57170507-1 KIEZIK, MICHAL

MANAGEMENT STRATEGY AND

INDUSTRY EVOLUTION

C.E. PROF. SHIGERU, ASABA

D.E. PROF. KANNO, HIROSHI D.E. Prof. SHIMIZU, TAKUMI

Summary

Abstract

The following thesis focuses on the business approaches for developing a viable secure computing product, as well as a sustainable business, in the cloud computing space. This document is structured in such a way as to first give the historical overview of the underlying internet technology involved in the current ecommerce world, the markets around the technology as well as the markets as they are now. This is followed up with a brief on the two most impactful rulings in recent history about customer data, namely HIPAA in the US and GDPR in the EU, as well as an overview of the cloud computing market - the most exposed and the fastest growing sector of the

online economy; rounding out the brief is a discussion on Amazon Web Services with a bird's eye overview of the competition. The following chapter is dedicated to the failures of security, as well as the current methodology of protecting data in the cloud; this is followed by an introduction to homomorphic encryption which is the pivotal point of the thesis. The next chapter then goes in depth on how to apply the technology to operations in the medical field, as well as looks at selecting the appropriate price models; it then goes into the overview of the competition and directly into the various risks on the playing field and appropriate mitigation strategies that are to be implemented. The conclusion is an overview of exit strategies for a startup in the homomorphic encryption space and a contains a few comparisons to recent acquisitions by Microsoft.

Key points, Industry Overview

In its original form the internet has been designed without any built-in security, instead it was built for openness and resiliency. This approach worked well when it was a tool for university research or between implicitly safe systems, but this is not the climate that exists now - where data needs to be kept secret, and any compromises in its integrity can mean massive financial or reputation damage. The security products that are currently on the market have been adapted over time, however so did the need for protection. While in the past it was adequate to focus on protecting data as it moved, and more recently as it is stored 'in the cloud', this is completely insufficient for

processing the data in 3rd party data centers or AI models. This is underscored by the recent legal changes, the HIPAA and GDPR, where penalties are assigned for inadequate protection of consumer data. There is a deep need for new digital security products and the underlying technology that is the most promising is the titular homomorphic encryption.

Key points, Data Security in Business

The risk for not protecting data assets is quite great, and for the companies that do not maintain the highest standards of security it has drastic consequences. In this chapter the data breach cases of Marriott Intl. (Hospitality), Adult Friend Finder (Social App.), and Quest Diagnostic (Medical Provider) are discussed and a more in-depth look is taken at the current practices in medicine in general. This chapter also underscores the importance of data in medicine.

Key Points, Cybersecurity in Medicine Business

This chapter begins with a deep dive into one medical data project and its implications, as well as the possibilities for improvement through the applications of services outlined herein. It goes on to develop application cases, and dives into which pricing model is appropriate for the segment. It is quickly established that the subscription model has the best fit, and the revenue stream options are discussed. At this point the competition is also analyzed, as well as the risks that it poses - the risk from both the incumbent giants of data management as well as the newest startups in the encryption

space. The chapter rounds out by cross analyzing the current strengths and weaknesses of the business as it stands now, and the appropriate movements that can be undertaken depending on the situation

Key Points, Conclusion

As the business model is set for the startup sector, the conclusion goes into the exit point, and discusses some similar cases of back end developer acquisitions. A few parting thoughts on the importance of finding a buyer that would be willing to continue developing the technology further are attached as well.

WBS

BUSINESS DRIVERS OF INNOVATION IN ENCRYPTION

Use Cases for Homomorphic Encryption

57170507-1 KIEZIK, MICHAL

MANAGEMENT STRATEGY AND

INDUSTRY EVOLUTION

C.E. PROF. SHIGERU, ASABA

D.E. PROF. KANNO, HIROSHI D.E. Prof. SHIMIZU, TAKUMI

Acknowledgements

I would first like to thank my thesis advisor Professor Shigeru Asaba of the Business School at Waseda University. Professor Asaba was always open to question about my research or writing while allowing me to find my own way while writing this paper - yet steering me in the right the direction when I would drift off topic.

I would also like to thank the staff in the Waseda Business School office for all their help during my time in the program, as well as my classmates for their encouragement and valuable discussions regarding research approaches. As the saying goes, "it takes a village" - and this is true of all works in life.

Additionally, I would also like to acknowledge Dr. Roupchan Hardowar, Dr. Tra Vu, and Rafal Sitkowski for always being available for consultation even when busy, and for always providing invaluable advice - especially the advice that helped me stay on target.

Finally, I must express my very profound gratitude to my mother and my sister for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Thank you all.

Table of Contents

CHAPTER 1. INTRODUCTION & RATIONALE

In the 21st century world, both the businesses and customers are entangled within the data they consume, and generate. Historically the handling of this business data has changed quite rapidly, going from the widespread local storage of paper based records as little as 50 years ago, through pockets of digital data within organizations, to the fully digital transmission and storage that is so common this day and age. In the more recent years the trend in shifting to treating data as a commodity has been prevalent, and beyond this the near-future trends are to treat data as a service - and this is already affecting the way we do business worldwide. This paper will explore the more recent move to the purely cloud based models that drive such trend shifts, the AI applications that push businesses to treat data as a service, as well as it's risks to businesses and customers that embrace these models. It will also offer insights into the technologies that underlie the critical components of the business models of the current dominant players as well as the venues of securing the data-in-use versus the current data-at-rest security models. We will also touch on the market development that these technologies are driving, and finally discuss the business use cases of the upcoming technological changes and possible long-term impacts of the technology trends.

CHAPTER 2. INDUSTRY OVERVIEW

Section 1. HISTORICAL BACKGROUND

2.1.1. ARPANET & Internet

Historically the internet was a tool designed for operations in trusted environments, and indeed "[i]n the 1960s, the Advanced Research Projects Agency (ARPA) of the Department of Defense began research of fundamental importance to the development and testing of computer communications networks." (Hauber, 1997, p.31) According to Hauber in the same chapter, ARPANET has by 1983 adopted the TCP/IP protocol that is in use until this day, and in mid 1980's has morphed completely into the Internet that we use now at the beginning of the 21st century. The original research and secure-by-default access to the network implied no need to further protect the communication, in-fact in the original TCP standard outlined in RFC 793 the following assumption is present:

> We assume that the local TCP is aware of the identity of the processes it serves and will check the authority of the process to use the connection specified. (Postel RFC 793, 1981)

Such a presumption of security and authorization has carried into the modern implementation of the underlying networks, and is only recently being addressed as we rely more on the networks for our daily lives. This point is driven further in the paper *"The effect of encryption on Internet purchase intent in multiple vendor and product risk settings"* whereas the author points out that:

> "The Internet was initially created to share information and not to support business process; thus few, if any, security mechanisms were implicitly included. Additionally, since the Internet is global, few legal protections exist for consumers." (Mascha, et. al. 2011, 401)

The essence of which is that the internet as a medium does not provide explicit or implicit security features or controls, and relies entirely on third party methods to protect the data that is passed through it.

2.1.2. Information-in-transit security

The first standard to be established to protect data exchange between non-government entities have culminated in the establishment and standardization of the OpenPGP encryption format for messages, as applied in message exchanges seen mainly through email, defined in August of 1998 in the following RFC standard:

> "Open-PGP software uses a combination of strong public-key and symmetric cryptography to provide security services for electronic communications and data storage. These services include confidentiality, key management, authentication, and digital signatures." (Callas, et. al. RFC 2440, 1998)

This methodology however focused only on protecting specific messages between specific users and required the use of third party software; moreover, it required the explicit and conscious decision to encrypt the message by the user, and required additional steps from the recipient side to read. It does not protect system level messages or any control data that would be needed for seamless commercial transactions, or for any use in the internet-of-things (IoT). Of particular note for the time period is the relative humor with which the latter, specifically IoT, was referred to - with an example given to standard published on April 1st, 1998, traditionally a tongue-in-cheek RFC publishing date, with respect to operations of remote controlled coffee machines:

> "Increasingly, home and consumer devices are being connected to the Internet. Early networking experiments demonstrated vending devices connected to the Internet for status monitoring [COKE]. One of the first remotely _operated_ machine to be hooked up to the Internet, the Internet Toaster, (controlled via SNMP) was debuted in 1990 [RFC 2235]." (Masinter RFC 2324, 1998)

While definitely possible as referred above attempt, the concept was not seriously considered due to technological limitations of the time - due to bandwidth limitations as well as a relative lack of stable network connectivity as the average household made due with a dialup connection; given this climate the security of the data transmitted as well as the potential uses of networks were left largely up to individual users. The change in approach to data security has taken a decade to be standardized in RFC 5246, describing the Transport Layer Security Protocol as per below:

> "The TLS protocol provides communications security over the Internet. The protocol allows client/server applications to communicate in a way that is designed to prevent eavesdropping tampering, or message forgery." (Dierks and Rescorla RFC 5246, 2008)

This development has ensured the adoption of link security in all browsers and various other services on the internet. However, the general storage of data on the far-end servers, otherwise known as the cloud has not been encompassed by this protection standard. And indeed, it has taken another 6 years for a standards advisory, not a requirement in itself, on the effects of monitoring of transmitted data to be published under RFC 7258, which defines the problem of pervasive internet data monitoring:

> "Pervasive Monitoring (PM) is widespread (and often covert) surveillance through intrusive gathering of protocol artefacts, including application content, or protocol metadata such as headers. [...] PM is distinguished by being indiscriminate and very large scale, rather than by introducing new types of technical compromise." (Farrell and Tschofenig RFC 7258, 2014)

The unfortunate reality is that for all transactions performed on the internet, there is likely a consistent leakage of semi-private data, however to date if needed the connection itself can be secured.

2.1.3. Information-at-rest security

While in recent years the security of data that is being transmitted has improved greatly, the security of the data that is stored at a remote location is still potentially weak. Additionally, all the large cloud data storage providers such as Amazon, Microsoft, etc. all state the highest standards of security and post the details of the methodology of data protection that is employed in their cloud services.

However, while the data is encrypted and stored as such on the far end servers there is no certainty of the privacy of any information actually stored in the cloud, as stated for example in the Google Cloud security document:

> " *The content contained herein is correct as of April 2017, and represents the status quo as of the time it was written. Google Cloud Platform's security policies and systems may change going forward, as we continually improve protection for our customers." (Google Cloud, 2017)*

Google's competitors provide equivalent functionality. Amazon provides a similar feature set in AWS in their various offerings such as Amazon Dynamo DB - per the service feature set description:

> "Amazon DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. [...] Also, DynamoDB offers encryption at rest, which eliminates the operational burden and complexity involved in protecting sensitive data. "(AWS, 2018)

It is worth noting however that while this capability is available, the overview of the limitations is limited and as always subject to change along with any policies. It is also not available in the region of China according to the same documentation set.

2.1.4. Information-in-use security

While over the last three decades the security of data created, data-in-transit, and data-at-rest has steadily increased as per the above examples, more often software is being treated as a service, and this will become more of a focus. This will be driven by the concept of AI as a Service, as mentioned in Forbes, the development over time has been as follows:

> "The first wave of cloud computing is attributed to platforms. […] The next big thing in the cloud was Infrastructure as a Service where customers could provision virtual machines and storage all by themselves. The third wave of cloud was centered around data. […] The next wave that would drive the growth of public cloud is artificial intelligence. Cloud providers are gearing up to offer a comprehensive stack that delivers AI as a Service." (Janakiram, 2018)

The article clarifies that the main driver behind this evolution of service offerings is the scale of data involved and the processing requirements at each point in ramp-up. According to Janakiram what further complicates the deployment of AI on-site is the amount of data that is required to create the AI model itself, and the near necessity of opening up the platform to other sources of said data than only the originating organization or business. This underlines an important need however, as in most cases for data to be processed it requires prior decryption, and it is therefore exposed to any attackers that may compromise the said system; additional methodology is required to protect data-in-use. A summary of the above schemes can be seen in Figure 1.

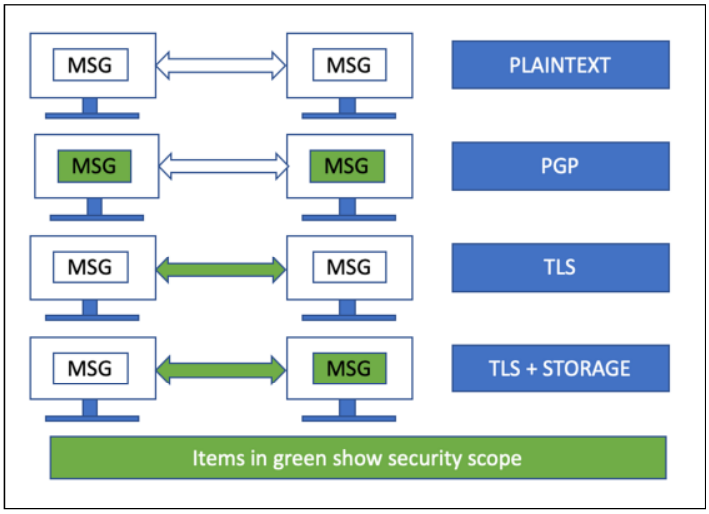


Figure 1 Summary of data protection scopes as outline above (Author, 2019)

2.1.5. Homomorphic encryption

To approach the task of protecting data-in-use, standard encryption schemes cannot be applied, as any data encrypted and stored, either on a far-end server or locally, is completely unreadable for as long as it stays encrypted. However other approaches to encryption can be used to circumvent this limitation and

secure data-in-use. According This is further elucidated according to an article published in the NTT Technical Journal regarding homomorphic encryption applications:

> "[I]f one uses traditional encryption, it also becomes impossible for the cloud operators to carry out any kind of processing of that data (even searching it, for example), as they would have to decrypt it first. This defeats the purpose of most applications of cloud computing. Fortunately, fully homomorphic encryption eliminates that limitation." (Tibouchi, 2014)

According to Tibouchi the prime candidate for homomorphic encryption methodologies application is indeed outsourced computing and, by virtue of utilizing it as supporting technology, it applies to cloud storage methodology as well. The general operation flow is demonstrated as it appears in the NTT Technical Journal in Figure 2.

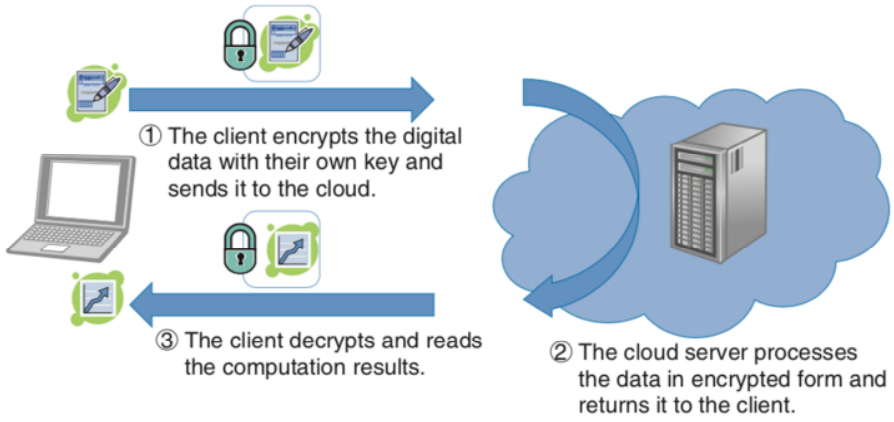


Figure 2 Outsourced Computing with Homomorphic Encryption (Source: Tibouchi)

Additional applications mentioned in the original article were the secure merger of data between different sources, without each source being able to decrypt the data; this results in a secure submission system, which becomes searchable and operable-on without cross leakage of data. The data is unavailable unless the private key to decrypt it is obtained.

According to Martins in his article "A Survey of Homomorphic Encryption: An Engineer Perspective", the homomorphic encryption operation:

> "[...]can be metaphorically explained by the jewelry shop problem (Gentry 2010), whose solution is represented in Figure [3]. A piece of gold is locked inside a glovebox, so that a worker may transform it into a ring. The ring is later removed when the glovebox is unlocked. "(Martins, 2017)

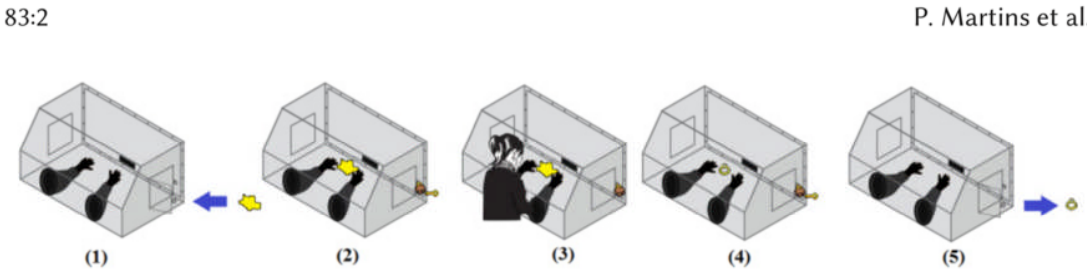


Figure 3 The Jewelry Shop Problem (Martens 2017)

Section 2. MARKET STRUCTURE OVERVIEW

The market outlook for cloud computing according to M2 Presswire is growth from USD 271.96 Billion in 2018 to USD 623.96 billion by 2023, with CAGR at 18.1%. They attribute this to the increased data generation via mobile apps and by the focus on delivering customized applications and experiences - all of which will require the storage and processing of larger amounts of personalized data, additional drivers are cost cutting measures shifting expenditure from CAPEX to OPEX due to the flexibility of the latter. Additionally, according to the same report, the entrance and proliferation of

machine learning and AI technologies, all of which depend on high volume of data to operate correctly, will drive growth at the later stages of the expansion. (M2 Presswire 2019). Meanwhile in the AI-as-a-Service sector is set to grow at a fairly aggressive CAGR of 45.2% between the same time span of 2018-2024, growing from USD 1.06 billion to 14.71 billion, according to Business Insights (BW Online Bureau 2019). The driving incentives is the use of third-party AI models with relatively small up-front investment and smaller data package requirements. However, the relative fragmentation of the provider market and the lack of skilled labor will tamp down the growth rates. This presents a relative opportunity to AIaaS providers in the long term as the market is flexible and a firm demand appears to exist.

Of special interest is the Healthcare Cloud Computing market which is set to grow at 22% CAGR to approximately USD 74.4 billion, according to Business Wire (Business Wire, 2019), and is set to be 65% of the SaaS market share worldwide.

Section 3. LEGAL CLIMATE

The current key pressure drivers for data protection in the world are the legal frameworks for medical data privacy in the US and general digital data in the EU, the HIPAA and GDPR respectively. Both of these regulatory frameworks lay out strict rules for the protection of customer data, as well as the handling of said data and the definition of security measures that apply therein. It is also important to note that both the frameworks provide for steep financial penalties for failing to uphold the safety of said customer data.

In the case of HIPAA, the key points of the framework with regards to customer data are as follows:

"(1) Ensure the confidentiality, integrity, and availability of all electronic protected health information the covered entity or business associate creates, receives, maintains, or transmits.

(2) Protect against any reasonably anticipated threats or hazards to the security or integrity of such information.

(3) Protect against any reasonably anticipated uses or disclosures of such information that are not permitted or required under subpart E of this part.

(4) Ensure compliance with this subpart by its workforce. "

(U.S. Dept. of Health and Human Services, 2013)

Additionally, HIPAA provides strict guidelines on the privacy of the patient, so that without explicit consent and need the patient data cannot be used across providers; this becomes especially cumbersome when considering building models of medical AI to serve diagnostic functions since as mentioned previously an AI model requires vast amounts of data to become accurate.

The second major legislation is the GDPR and it is similarly restrictive, however in a far broader scope; what is of specific interest here is the regulation what constitutes high risk activities, whereas:

" Evaluation or scoring, including profiling and predicting, especially from 'aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behavior, location or movements' (recitals 71 and 91)." (EDPB, 2017)

This broad statement encompasses essentially any items that involve machine learning or AI processes; furthermore, the penalties for certain failure scenarios in case of high risk activities ranges from 10 million Euro, to 2% of the total worldwide annual turnover, according to the same publication - setting a heavy incentive for the providers to protect the customer data to the best of their ability, but also to avoid the so called high risk scenarios. This unless approached from a highly secure standpoint could potentially serve to discourage the growth of AI decision making and the availability of data for machine learning.

Section 4. CLOUD COMPUTING

An important component of all business that is internet enabled in this day and age is the concept of cloud computing, and transaction processing. This is not only crucial to enabling ecommerce but also to cutting costs that are incurred by organizations via drastically reducing capital expenditures, and if utilized correctly limiting operational expenditures. This is largely due to being able to take advantage of economies of scale enjoyed by the cloud service providers; economies of scale that many companies would not be able to enjoy - especially companies that do not deeply reach into the computing space as a business, such as manufacturing, service, banking or medical industries. And while the line between computing technology companies and these industries blurs more every day, it still is up to those specialized cloud computing aggregators to provide efficient back office services such as storage and computing power, as well as network effect services such as payment and transactions. To touch on the importance of these services, we can take a look at Amazon Web Services as the largest player in the cloud computing market.

2.4.1. Amazon Web Services

Amazon Web Services, commonly known as AWS, has begun operations in 2006 as an offshoot of Amazon; given that Amazon as an ecommerce platform had fairly heavy demands for IT infrastructure, the leveraging of that infrastructure to start offering services to other firms was a possibility. According to the corporate website, the main target niche of AWS is companies that are looking for:

> "[...]the opportunity to replace up-front capital infrastructure expenses with low variable costs that scale with your business. With the Cloud, businesses no longer need to plan for and procure servers and other IT infrastructure weeks or months in advance. Instead, they can instantly spin up hundreds or thousands of servers in minutes and deliver results faster." (Amazon, 2019)

This approach at offering services has paid off, especially in recent years as reported by the Amazon Press Releases, whereas the fourth quarter 2018 sales have been up by 20%, sitting at USD 72.4 billion.

When compared against 2017, the performance was staggeringly positive with key metrics being, as self-reported:

> "Operating cash flow increased 67% to $30.7 billion for the trailing twelve months, compared with $18.4 billion for the trailing twelve months ended December 31, 2017. Free cash flow increased to $19.4 billion for the trailing twelve months, compared with $8.3 billion for the trailing twelve months ended December 31, 2017. " (Amazon, 2019)

This seems to be backed up by market research for first quarter of 2019, with Canalys reporting 41% growth in first quarter; and while this is approximately half the growth when compared to Microsoft and Google, Amazon is heavily ahead of their competitors in size, as can be seen in the Canalys report chart in figure 4. Also, according to Canalys, AWS will soon start shipping devices to customer datacenters to closer integrate with any on-site networks, and allow for further hybridizations with customer data centers - increasing their value offering and entrenching themselves in those relationships even deeper. (Canalys, 2019)

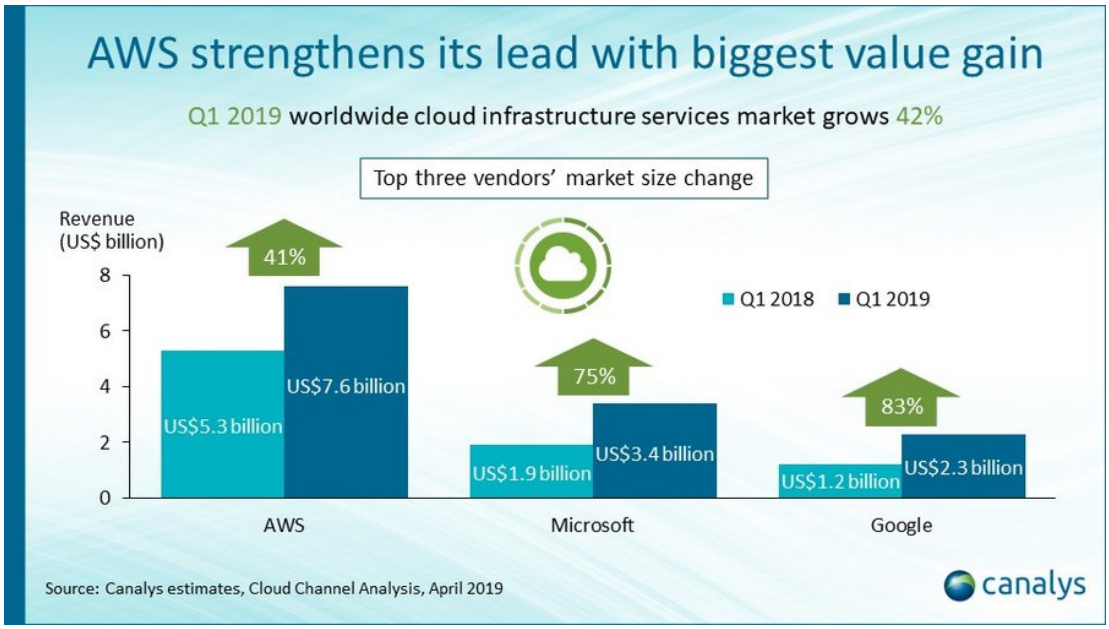


Figure 4 Growth Comparison with Other Major Cloud Providers (Canalys 2019)

For the overall market growth perspective, we can refer to the Business Insider article regarding cloud computing business from last year, that notes the cloud computing market is still in its early stages; and points to largely 4 players that will continue to control the entire cloud computing market - namely Amazon, Microsoft, Google, and AliBaba. The overall market is said to be competing with their customers own in-house systems spending - Business Insiders' Wolverton and Lee note:

> "As cloud spending expands, it does so in part by eating into money companies used to spend on such things as maintaining and running applications on their own servers. Last year, spending on cloud services accounted for about 8% of the total potential market. That should jump to about 15% by 2021, Goldman Sachs estimated. In other words, a growing portion of corporate IT budgets is likely to go to the cloud." (Wolverton and Lee 2018)

This is well visualized in the Goldman Sachs market overview referenced in the article shown in Figure 5. Of particular note is the rapid decline of the 'other' category, indicating very strong consolidation of the entire market between the four major players.

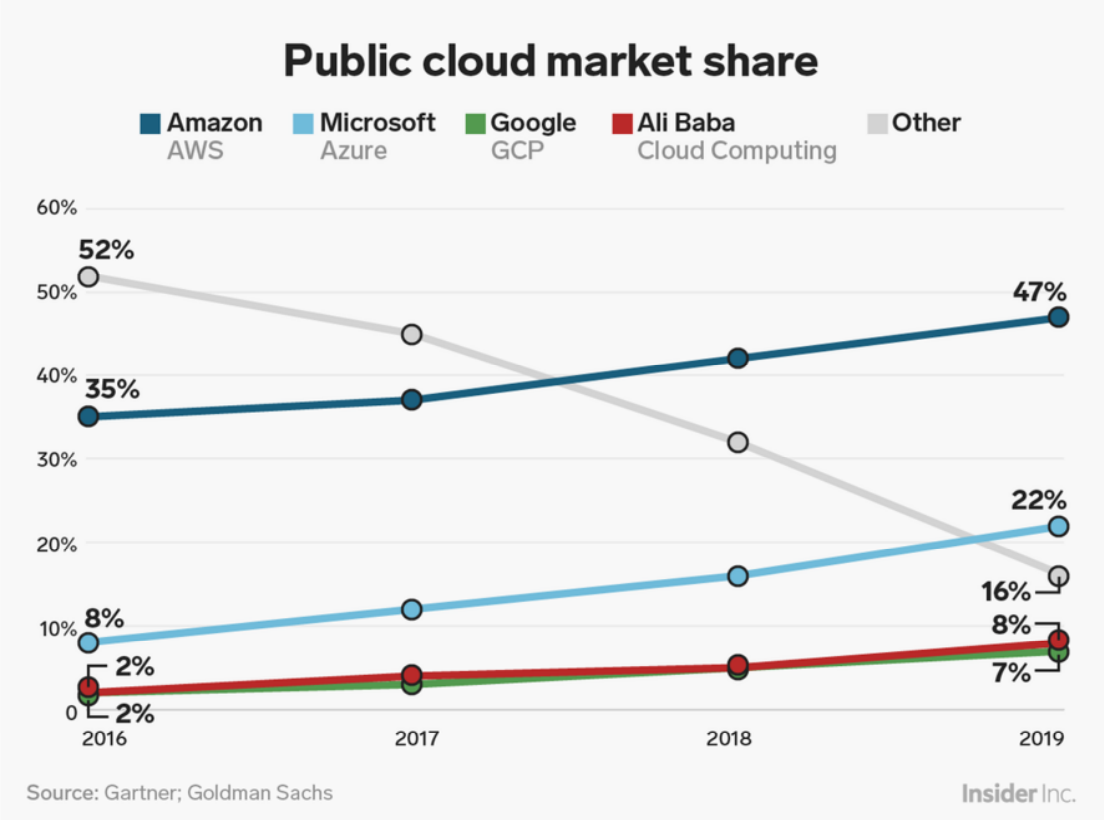


Figure 5 Market Size and Consolidation (Business Insider 2018)

Amazon is largely successful due to its scale, and the powerful inertia behind its growth; however, it also holds a policy of strategic acquisitions of technology as is mentioned in its governance policies. According to its internal documents it holds targeted investments towards growth as one of the core tenets of business, as stated below:

> Make bold investment decisions in light of long-term leadership considerations rather than short-term profitability considerations. There is more innovation ahead of us than behind us, and to that end, we are committed to extending our leadership in e-commerce in a way that benefits customers and therefore, inherently, investors -- you can't do one without the other. Some of these bold investments will pay off, others will not, but we will have learned a valuable lesson in either case. (Amazon 2019)

To that end, the two notable acquisitions relating to its success in the cloud are CloudEndure, which focuses on disaster resiliency, as per their company page:

> CloudEndure offers highly automated disaster recovery and migration solutions into AWS. With support for any source infrastructure and all applications running on supported operating systems, CloudEndure ensures that your entire IT landscape will remain robust and reliable as it continues to grow. (CloudEndure, 2019)

As well as TSO Logic, which focuses on planning capacity and costs estimation for the customers; as per their company page:

> We show the actual performance, utilization and TCO of our customers' current environments and cloud alternatives, so they can optimize their spending and develop a data-backed business case for change. (TSO Logic, 2019)

The overall investigation into the Amazon family regarding advanced encryption and data protection technologies yielded information on an apparent partnership with Gemalto, by offering encryption of storage as noted in the AWS marketplace article on data protection:

> "Gemalto SafeNet ProtectV for Amazon EC2 (EC2) secures sensitive and highly-regulated data by encrypting entire virtual machine instances and attached storage volumes." (Amazon, 2019)

Gemalto is at its heart an encryption key management company, that operates according to its company page:

> "by combining Digital Identity and Data Protection across the entire digital service lifecycle. " (Gemalto, 2019).

While strictly a higher level of security than simply encrypting data with localy stored keys, it still requires data to be decrypted while in use.

The above Amazon case serves to highlights a few important points regarding the cyber security segment, as it concerns the cloud computing world:

- An opportunity exists for cooperating with large cloud computing providers on security and data protection
- Advanced encryption avenues are not exhausted, and in-fact are just starting to develop
- The niche of encrypting data in use remains unfilled

This points to an optimistic outlook for cyber security startups, whether they aim to cooperate with the cloud computing provider giants, or if they approach the customers of those providers directly and offer encryption and data security services that would be transparent to the underlying infrastructure.

CHAPTER 3. DATA SECURITY IN BUSINESS

Section 1. DATA BREACH INCIDENTS

3.1.1. Data Breaches

The simplest driving argument for the use of advanced encryption methodologies is to take an overview of some of the high-profile data breach cases that have occurred in the recent years, and the approaches that could be taken to mitigate the risk of recurrence by applying homomorphic encryption methodologies to the storage and processing of data. It is important to note that the data breach does not have to be completely prevented to secure the customer data, as long as the data is non-recoverable for the attacked; in this respect, we will consider any data that is unrecoverable or unreadable for the attacker to be protected. According to news sources, and listed in CSO Online by Taylor Amerding the top data breach events and the number of customers affected until end of year 2018 are as follows (Amerding, 2018):

- Yahoo - 3 billion customers
- Mariott International - 500 million customers
- Adult Friend Finder - 412 million customers
- eBay - 145 million customers
- Equifax - 209 million customers
- Heartland Payment Systems - 134 million customers
- Target Stores - 110 million customers
- TJX Companies - 94 million customers
- Uber - 57 million customers
- JP Morgan Chase - 83 million customers

For the first half of 2019 we can refer to the ZDNet article to identify additional 5 mass data breach scenarios:

- First American - 885 million customers
- Facebook - 540 million customers
- Australian National University - 200 million
- Canva - 139 million customers
- Quest Diagnostics / AMCA - 11.9 million

In most of the above cases the data breach would not be as critical if the data obtained by the attackers was not linkable to the actual customer that the data describes. Considering the functionality provided by the above services as well as functionality of homomorphic encryption as applicable to these business models, we will go in depth on 3 of these cases below.

3.1.2. Marriott International

In November of 2018 a data breach occurred at Marriott International, with approximately half a billion victims according to Forbes magazine. The article mentions the method of discovering the breach was automatic, and that:

> "One of Accenture's products, IBM Guardium, had detected an anomaly on the Starwood guest reservation database on September 7. [...] On September 10, Marriott called on third-party investigators to look into whether it had been breached. Soon afterwards, malware on the Starwood IT systems was found: A Remote Access Trojan (RAT), which allows hackers to covertly access, surveil and gain control over a computer." (O'Flaherty 2019)

The key take-away here is that even though automatic detection systems have recognized the attack and theft of data, it took several days for the investigation to begin in earnest - by then the damage has been done. The data while surely stored in a secure environment, was still completely accessible when in-use -

and therefore available to those that had elevated access to the system. Considering that according to the same article the data stolen was passport data and guest reservation data, both not necessarily needed to be processed on the cloud side, using encryption schemes that would provide security in processing would have decreased the impact considerably.

3.1.3. Adult Friend Finder

In November 2016, the FriendFinder network was hacked and approximately 412 million user records were stolen according to the Forbes article on the matter. While the site was of reputation sensitive nature, it did not adhere to good standards of security. According to the same article:

> "Apart from the staggering volume of victims and the sensitive nature of the activity going on at AdultFriendFinder, there's another troubling detail about this hack. Much of the user data was stored as plain text. [...] Even customers who think they'd cut ties with AdultFriendFinder have been caught with their pants down. Deleted accounts were still listed among the active ones, they had merely been flagged." (Mathews 2016)

The scope of the breach could have been much smaller if the data was encrypted, as it would have been largely useless to hackers to obtain only login information - as in this case the most damaging aspects were the customers real life contact details being leaked.

3.1.4. Quest Diagnostics

Quest diagnostics breach resulted in the theft of approximately the data of 12 million customers according to Forbes. This has exposed Quest to a lawsuit seeking USD 5 million in damages. According to the same article:

> "Filings with the United States Securities and Exchange Commission made by Quest Diagnostics revealed that information including patient names, dates of birth, addresses, phone numbers, dates of service, care providers and account balances were exposed. In SEC filings made by

LabCorp, it was disclosed that about 200,000 of its customers had credit card or bank account information stolen. " (Dellinger 2019)

The same article points out the obvious problem of health care providers being especially vulnerable to breaches given the nature of the sensitive data that said providers store on their users, as well as the fact that health care providers tend not to focus on technology unless it is directly related to their function. This becomes more problematic if we consider the fact that health care is becoming more computerized every day, and that the application of AI in healthcare will result in the data being distributed to even more systems throughout the organization. Additional considerations will also have to be given to any data that is present, or is moved to, any cloud computing environments.

Section 2. ADVANCED ENCRYPTION APPLICATIONS IN MEDICINE

As the use of data in the medical world continues to grow, the concerns for the security and the misuse of the data continue to increase. As mentioned in the previous chapter, legislation to the effect of protecting patient data is consistently being tightened, and the penalties for potential data misuse and data loss are increasing. This alone is a potential driver for implementing advanced encryption models in the medical field. It should also be recognized that not only the protection of data from theft is needed but also a few other authenticity verifications services. According to Roy, et.al. the following are critical points when it comes to handling digital patient data:

- Maintaining the Privacy of the Individuals
- Maintaining the Reliability
- Maintaining the Authenticity
- Maintaining the Integrity

Based on these four tenets Roy, et. al. state:

> "The medical images often contain highly sensitive data that are directly useful for the diagnosis purpose. It is necessary to protect these data from unauthorized access. Someone can steal these data over the network in case of weak security. It can destroy the privacy of the patient. There will be high chance of misusing these data to gain some profit from it which can be very harmful for the patient as well as for the government. So it is one of the major point that is to be considered during the design of any security algorithms." (Roy, et. al. 2019)

While in this case the focus is medial images, all patient data has a similar potential for harm if not controlled.

However, above and beyond the compliance and security requirements is the functional driver of patients trusting their medical providers with their data - and that is the diagnosis and correction of any medical problems; this becomes more important as the world continues to drive towards AI and machine learning use and such data becomes absolutely necessary for the treatment of any diseases or conditions. Additionally, the data sources of patient data continue increasing, especially with the continued introduction of medical monitoring devices and personal health monitoring devices. The main rationale for continuous monitoring of health devices is outlined by Masood et. al. in their overview of data security in healthcare:

> "Pervasive healthcare monitoring provides rich contextual information to handle the odd conditions of chronically ill patients. Constant monitoring and an early medical response not only increase the life quality of elderly and chronically ill people but also help families and parents by providing high-quality healthcare to their young babies and paralyzed children." (Masood, et. al. 2018)

In the above context, it is quite advantageous to the care service provider to host and combine the data that is generated by the medical devices in the cloud, as access to and from the devices may vary, and the

patients may or may not be on a hospital control network; this is demonstrated by the simple flow diagram from the same article as above in Figure 6.

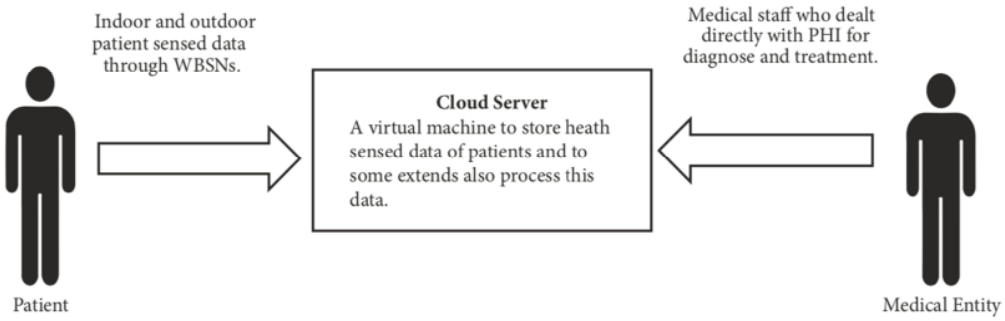


Figure 6 Data Exchange and Collection between Patients and Medical Entities (Masood et. al. 2018)

This data storage methodology however exposes the data to additional breach risk, as well as potential changes to the terms of use of the cloud service provider. Additionally, the healthcare service providers can offer value added options as outlined by Masood et. al., in the following ecosystem diagram under Figure 7.

Figure 7 Healthcare Value Add Services (Masood, et. al. 2018)

The above sections can be broken down into a few main points of data collection and processing, based on data input, storage and decision making. The main components are as follow:

- User Interface / User Experience - the components that the user interacts with
- Medical devices - the components generating user data
- Diagnostics and Telemedicine - the components that act and utilize data

The various aspects of data collected as well as the processing of data linked to the user have need to be secured not only while the data is stored and transmitted, but also when the data is being processed. This is exponentially more important when considering automatic data diagnostics such as AI and machine learning algorithms. Therefore, it is relevant to gain some insight into customer behavior in relation to security measures employed by the service providers as well as the current approaches to managing patient data.

The converse scenario, that of protecting the AI models that are built based on the medical data of patients serviced by the provider must also be considered; given that the creation of AI and machine learning modes requires large amount of carefully adjusted data, the AI model must be exposed to patient data from various sources - however given that patient data is sensitive and cannot be easily traded between organizations the sourcing of data is restricted. On the other hand, as AI models are a potential source of revenue for care providers, as well as the analysis of such models could expose the providers to venues of attack, AI models are to be considered sensitive intellectual property and must be protected with the same degree of care as the patient data that they are built on.

3.2.1. Current approaches

Current approaches to protecting patient data rely mainly on the on-site network protection when the data is stored by the provider; however, such approaches limit the potential application of patient data, as well as decrease portability and preclude the use of the data in AI model creation. One potential way of circumventing such limitations is the use of federated models for data exchange between organizations; in which the machine learning model itself if shared and patient data is fed into

the system on a one-by-one basis, or by strict control of access to the models themselves. While this does not remove the risk of data breaches, and indeed given the distributed form of the federated network it does increase the attack surface exposure, it does considerably limit the exposure from a one-time breach to whatever data source or service is breached. Additionally, the question of protecting individual datum remains unresolved in this model, each datum must be encrypted separately, and in case of AI models the model needs to be shared and retrieved - exposing it to investigation and potential loss or dilution of IP. The flaw in this approach is pointed out in a design research article considering these quandaries as far back as 2004, where data volumes were orders of magnitude smaller than currently:

> "Our university's radiology department conducts some 380,000 examinations and produces 9 TB of digital data annually. Images are initially collected and stored in the industry standard DICOM (Digital Imaging and Communications in Medicine) format. HIPAA will require that these images be stored and communicated in encrypted form. [...]What is the impact on the radiologist's workflow if every image has to be encrypted before storage and decrypted before viewing? " (Weaver, et.al. 2003)

Considering the above, it is crucial to provide both the service providers and the patients with the ability to securely store and manage the data that is generated and collated about health. It is also quite self-evident where the advantages of machine learning are in these scenarios, as processing vast quantities of data is exactly the use case that AI can be applied to best. It is also evident that such vast quantities of data are beneficial to the organizations that can create their AI models as they can pivot those data stores into services where the AI models they create can be used by other medical care providers. This relates back to the previous chapter on AI as a Service, whereas the providers could expose their AI constructs to the outside world and allow smaller organizations to take advantage of their expertise, all without exposing patient data or any information that is related to the current customer set. This can only be accomplished however if the data is handled securely, and it directly relates to the purpose of this paper;

namely the look at business models in securing data-in-use by application of homomorphic encryption technologies.

It is worth noting that another model for AI learning exists that distributes the AI model itself to various processing nodes; this approach is called Federated Learning and the security inherent in it would be based on outsourcing the AI model itself to the data holder, instead of the other way around. The operation of the model is summarized by Bhattacharya in the article 'The Dawn of AI: Federated Learning" as below, with specific emphasis on the fact that the original data always remains on the customer end, and thus never needs to be protected on the cloud:

> "Functionally, a mobile device that is a part of a FL computing architecture, downloads a model that is meant for running on mobile devices. It then runs the model locally on the phone and improves it by learning from data stored there. Subsequently, it summarizes the changes as a small update, typically containing the model parameters and corresponding weights. [...] Most importantly, all the training data remains on user's device, and no individual updates are identifiably stored in the cloud."

The general flow of learning data can be seen in figure 8, whereas the customer data never leaves the customer device, instead only processing the data using the AI model provided, and returns the modified learning data.

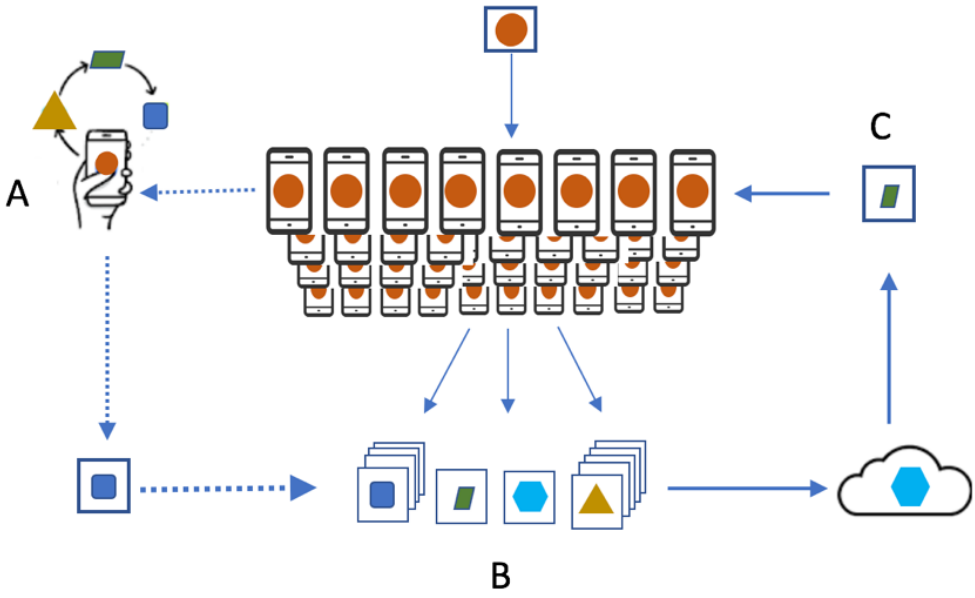


Figure 8 Federated AI Model Learning Avoids Cloud Storage of Client Data (Bhattacharya 2019)

This approach, while allowing the customer data to remain within their control, makes an important tradeoff as it provides the AI model to the customer device. This in turn exposes the provider's AI model to the world instead of the customer data; in cases of providers that have invested considerable resources in building said AI models, or where the AI models are the source of competitive advantage, this could prove highly damaging to the provider's business model as the said AI model leaves their sphere of control. Once the AI model is loaded onto the client device or environment it can be copied or reverse engineered and any such advantage could be lost. On the other hand, given that customer devices are getting more powerful every day and not taking advantage of these resources will be quite wasteful - however to properly operate such networks the cybersecurity will have to then focus on the AI components. Bhattachaya predicts that:

"In the next few years, model building and computation on the edge, based on Federated Learning and secured with Homomorphic Encryption will make a significant progress. [...]Distributing the heavy duty analytics and computations over smartphones “on the edge”, as opposed to central computing facilities, will drastically reduce time to develop data products such as hyper-personalized recommendation engines, e-commerce pricing engines etc."

(Bhattachaya 2019)

CHAPTER 4. CYBERSECURITY IN MEDICINE BUSINESS

Section 1. APPLICATION METHODS

The key to developing a business around homomorphic applications in the medical field is not to develop additional uses for homomorphic encryption within the health care space but instead integrate it into current and near-future applications. As demonstrated in the previous chapters, data security is both critical in any applications where customer data is stored, but it is also incredibly difficult to do so. It is also clear that as customer data storage and processing moves into the cloud this will become even more so problematic as long as data needs to be decrypted before it is used. However properly developed applications of homomorphic encryption solves a large portion of this problem, as data can be stored and completely securely on the cloud without the means to decrypt it being available to any attackers even if the system is compromised. Therefore, the business focus here is to provide the underlying technology as well as expertise to integrate the said technology into the research models as needed; this can be done in three stages of involvement, that would have to be re-evaluated at every gate over time evaluate the next steps. However, given the development of computing and storage technologies the focus of development over the long term should be to platformize the offering and the alternative opportunity is to offer the solution in SaaS model. The general time flow overview of development is shown in Figure 9 below.

Figure 9 Development Focus in Homomorphic Encryption for Storage and AI (Author 2019)

Section 2. GIFT PROJECT CASE STUDY

4.2.1. Background

The GIFT platform is designed to provide patient data, care information, as well as supporting services for fetal surgery. The rationale of the GIFT platform is outlined in the paper by Doel, et.al. as below:

> Clinical imaging data are essential for developing research software for computer- aided diagnosis, treatment planning and image-guided surgery, yet existing systems are poorly suited for data sharing between healthcare and academia: research systems rarely provide an integrated approach for data exchange with clinicians; hospital systems are focused towards clinical patient care with limited access for external researchers; and safe haven environments are not well suited to algorithm development. (Doel, 2016)

Given that medical imaging is widely used in clinical practice, the continuing development of novel imaging modalities, improved protocols and computer-assisted analysis software has huge potential to

improve disease assessment and patient outcome. These advances require ever closer collaboration between academia, industry and healthcare providers. However, given that the designers of the GIFT system were constrained by privacy considerations, the system itself is designed to anonymize the data that it stores and operates on. This tends to limit the usefulness of the data set provided, especially in the case of machine learning applications, but even in the case of retrieving a specific data set that may be needed for surgery of the patient in question. It becomes essentially a one-way system to creating and storing research data. The operational diagram is taken from the same article and show in Figure 10.

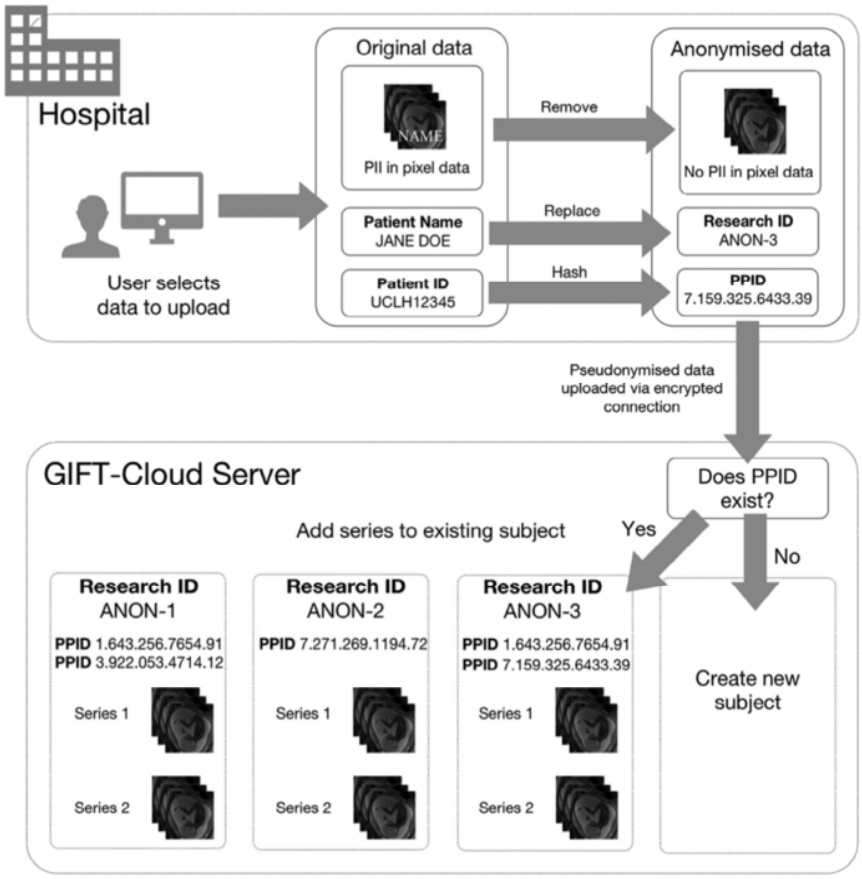


Figure 10 Implementation of GIFT Data Storage (Doel and Shakir 2016)

However, if homomorphic encryption solutions are applied to the above project, several avenues of research as well as diagnostics open up. This would allow the operating organizations hosting the GIFT system to not only gather the data that is available within their organizations, but also allow the submission of data from far smaller entities, as well as leverage the data available to train AI and machine learning constructs. This in turn would allow the hosting organizations to monetize the service without risking the exposure of patient data. The potential workflow for implementing such services is shown in Figure 11 below.

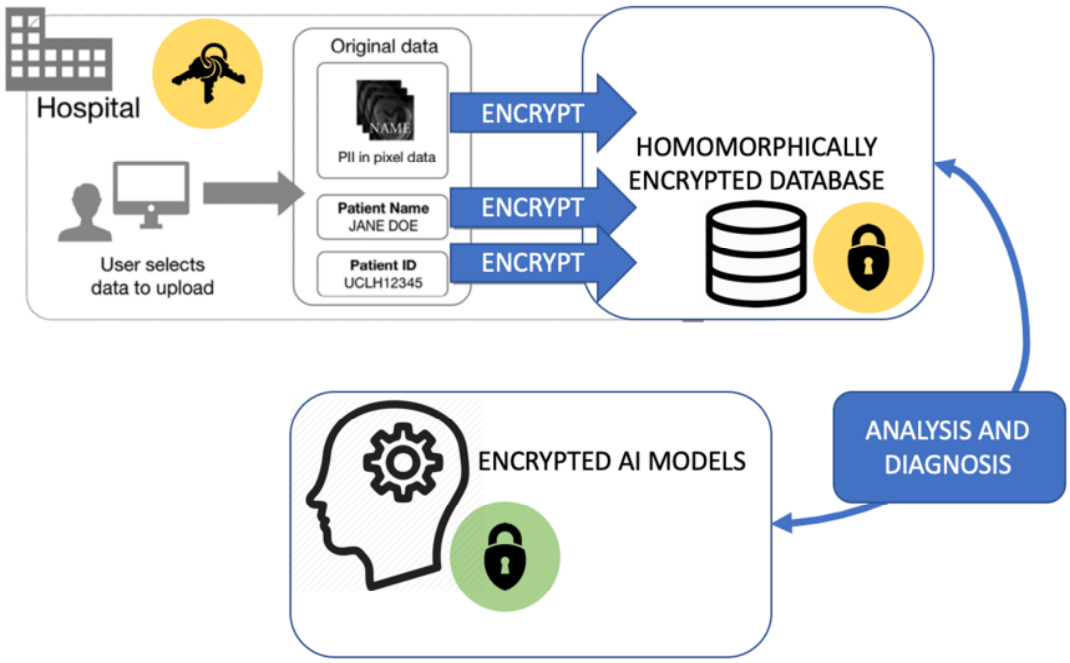


Figure 11 Potential Implementation of Homomorphic Encryption with GIFT (Author 2019)

The above approach would be compatible with implementing the homomorphic encryption schemes as a stand-alone software implementation but would greatly benefit in Software-as-a-Service frameworks as well, allowing for flexible pricing. The additional benefit of securing the data at the very beginning of

the process would also allow the healthcare provider, as well as the software provider, to leverage cloud computing to the fullest - both in storage of patient data, as well as in processing of said data. This could potentially decrease the liability of the providers under HIPAA rules as well, given that it would clearly meet the standards of data protection. Given that the keys to the data encryption and decryptions would be held at the healthcare provider side, it would also be possible to retrieve the patient data in case it was necessary - such as in case of required surgery, or follow up diagnosis.

4.2.2. Approach & Pricing

The approach to operating a business that provides homomorphic encryption technology depends heavily on the stage of software evolution the firm currently holds as their intellectual property as mentioned and demonstrated in Figure 9 previously. The general approaches would be as follows:

- At the development stage a consulting approach should be adopted, especially since the company is in the research and development stage of the software and product creation. This, depending on market conditions will require several proof-of-concept projects; given the relative obscurity of the technology aggressive rates would be necessary for the consulting portion of the business with less than the standard times-3 multiplier; optimally a times-2 multiplier for cost per-headcount would allow for market penetration and continuing development effort
- At the ready-to-market stage where a product is available and offered the approach shall shift from offering consulting based projects to implementation of the software as an 'off-the-shelf' solution. This would enable the implementation of the system on the customer side to shift from the relatively high cost resources such as developers, to potentially cheaper integrator resources, saving costs and freeing up the development team to continue creating and improving the core products
- At the mature product stage the focus should shift to offering the solution as a Software-as-a-Service package; this would enable the licensing to be increasingly fine

grained and implementation could potentially shift completely into a self-service model; this is especially true if the SaaS approach can be deployed entirely into the cloud with the underlying infrastructure offered as a package deal

- At the final stage the option to continue expanding operations exits, this would require developing and managing an in-house platform version of the product - potentially acquiring the medical models and services into the greater SaaS whole, or develop a marketplace that allows for providers to offer their models to the greater ecosystem of users of the software product itself
- The potential also exists, especially at the later stages of the lifecycle, of an acquisition by one of the major cloud computing platforms as the encryption models do mesh well with the core business of most providers.

Further considering pricing approaches, especially given the constantly changing landscape of software pricing, can be referenced to several research works. We will use the research published by Arto Ojala in "Adjusting Software Revenue and Pricing Strategies in the Era of Cloud Computing". The essential findings from the research are as follows:

> "The findings indicate that servitization of the software offering makes it possible to adjust revenue and pricing strategies relative to market competition. Depending on the competitive situation in the market, firms apply mixed revenue models, or else a hybrid pricing mechanism, to protect their business against rivalry and substitutes. The software renting model has several advantages which significantly help software vendors to expand their business opportunities. However, in some cases, powerful customers are able to limit the revenue and pricing options." (Ojala, 2015)

Given this data we can use the decision tree chart that was published in the above article, to gain some insight on behavior on companies that are faced with this question, and highlight the strategy that would work the best in our chosen case of medical data security; the chart in Figure 12 has the overlay for the

relevant branches to our scenario, green being items that are in favor of the decision, and in red the items that are against the decision.

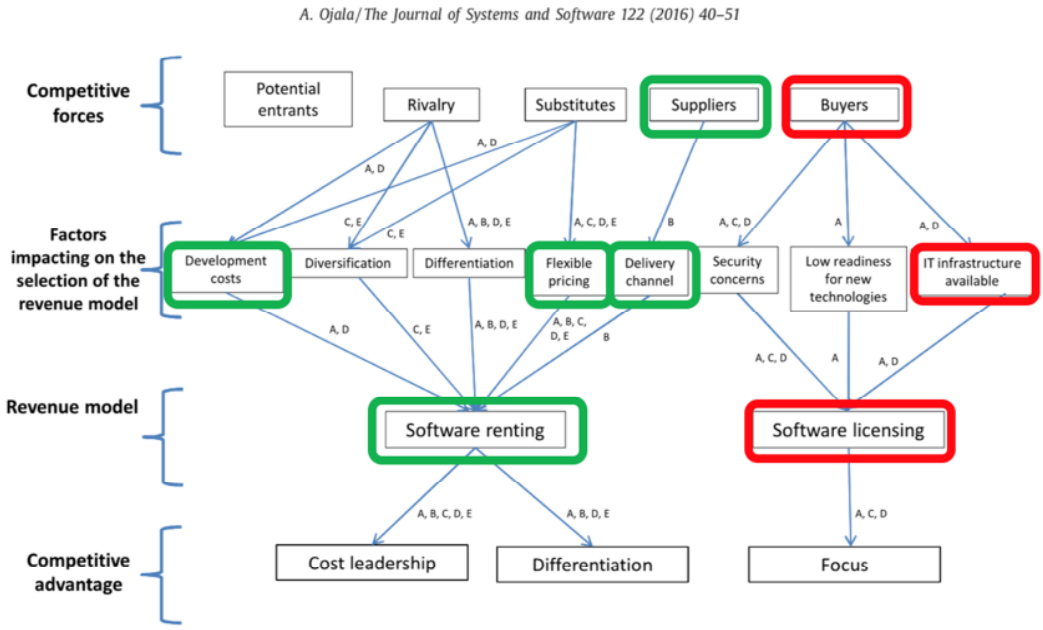


Figure 12 Suggested Pricing Options with Overlay (Ojala 2015)

The rationale for the above is as follows:

- Development costs are highlighted on the software provider side as these are continuous, as well as dependent on the rivalry and new entrants, additional costs may be incurred for hosting and processing, etc.
- Flexible pricing is highlighted given that additional costs and profit center is possible via customization of product, consulting with medical providers, etc.
- Delivery channel is highlighted due to targeting the in-cloud segment, as well as attempting to shift non-cloud customers into the cloud, additionally SaaS marketplaces could be leveraged depending on the target clients existing infrastructures

- IT Infrastructure highlighted as a negative given that medical services providers do not have information technology expertise as a core competency, therefore even in cases where there is an IT presence the customer may want to prevent the expansion of duties from existing staff and even potentially move away from in-house products

The above sets the strategy firmly in the 'software renting' model, which supports the rationale for Software-as-a-Service approach to development, as well as the platform targets for the far-end development of the company strategy.

Section 3. COMPETITIVE LANDSCAPE

While the competitive landscape of cloud computing providers is relatively saturated by four very large companies as mentioned in the previous chapter, the situation in the cyber security area is quite more fragmented, with a mix of large providers focusing mainly on incumbent technologies such as network protection, traditional encryption, antivirus, etc. Taking a look at Giarratana's publication, 'Research Policy 33', we can focus on the large companies that were operating in 1998, as in Figure 13 below.

Table 2
World market leaders in ESI at 1998

Rank	Firm	Revenues (US$ in million)	World market share	Year of entry	Firms in the same entry cohort
1	Network Ass.	990	0.171	1993	18
2	Symantec	578.4	0.099	1990	8
3	RSA Data Security	171.3	0.029	1991	17
4	Check Point	141.9	0.024	1995	35
5	Rainbow Tech.	109.2	0.018	1998	56
6	Axent Tech	101	0.017	1994	29
7	Trend Micro	86.2	0.014	1991	17
8	Secure Computing	61.4	0.010	1994	29
9	Entrust Tech.	49	0.008	1997	57
10	Cylink	42.8	0.007	1995	35
11	SystemSoft	42.6	0.007	1998	56
12	VeriSign	38.9	0.006	1998	56
13	BindView	38.5	0.006	1995	35
14	Aladdin	36.1	0.006	1997	57
15	Safenet	23.2	0.005	1998	56
Total		2487.3	0.429		

Source: IDC Corporation, Infotrac.

Figure 13 Startups in Encryption as of the End of 20th Century (Giarratana 2004)

Comparing these companies with their current status based on their revenue numbers we see considerable consolidation in Table 1 below:

Table 1 Revenue Comparison 1998-2018 (Owler 2019)

Name	Revenue 1998	Revenue 2018	Growth 20y	Notes
McAfee (Network As.)	0.99	2.1	212%	
Symantec	0.58	4.7	810%	
Checkpoint	0.171	1.4	819%	
Rainbow	0.142	0	0%	Acquired by Safenet
Axent	0.109	0	0%	Acquired by McAfee
Trend Micro	0.101	1.2	1188%	
Secure Computing	0.086	0	0%	Acquired by McAfee
Entrust Tech	0.061	0.59	967%	
Cylink	0.049	0	0%	N/A

SystemSoft	0.043	0	0%	N/A
Verisign	0.039	1.2	3077%	
BindView	0.039	0	0%	Acquired by Symantec
Aladdin	0.036	0	0%	N/A
Safenet	0.023	0	0%	Acquired by Gemalto
Source:	https://www.owler.com/			Rev. in Billion USD

Considering the above data however, it is worth noting that the largest companies are integrated providers that focus heavily on the consumer and office-space PC market, especially McAfee. Therefore, the competitive space can be considered as not extremely saturated when we look at the core competencies of the companies with regards to secure computing. It is also a positive outlook in terms of startup exit options as the larger firms in the space tend to acquire the companies that develop technologies outside of the core portfolio of the larger players. This can contribute to the net evaluation of exit options in terms of startup valuation.

If we instead look at the current competitive space of startups within the space of secure computing we will see the competitors map as below in Figure 14.

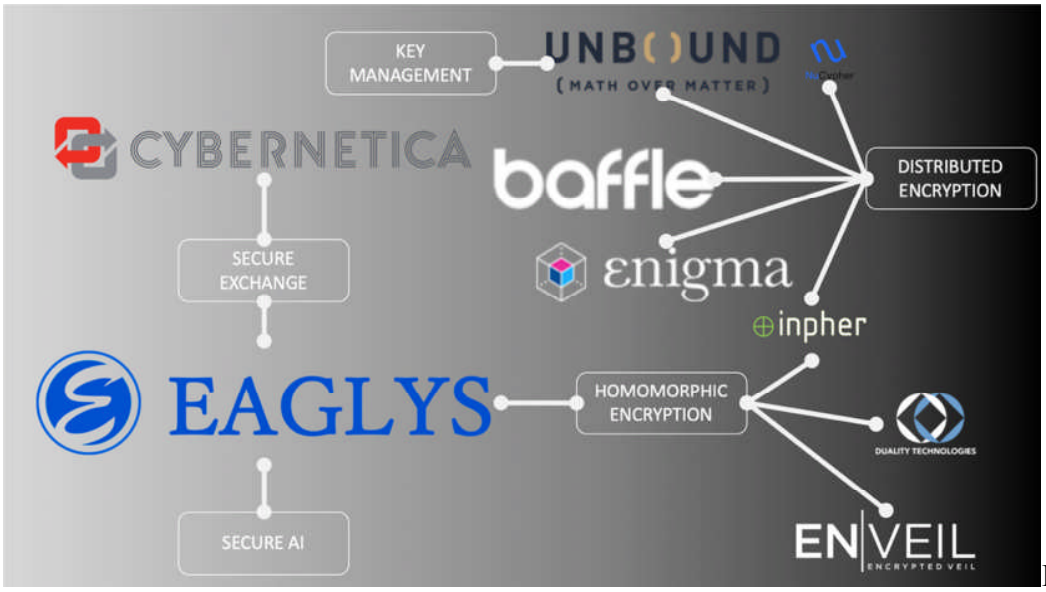


Figure 14 Competitor Map in Secure Computing (Author 2019) [all logos are the property of respective owners, information obtained from corporate pages]

Generally, all companies in the secure computing space are in the startup phase and as such the data resolution on the competitive landscape funding varies. That said the only company currently operating in both the homomorphic based secure encryption database space and the homomorphic based secure AI applications is Eaglys, Inc. with their offering of both homomorphicaly encrypted databases and homomorphicaly encrypted AI models. Given the startup mode of the company, we will look into the appropriate risks and mitigations in the next section. It is also worth noting that Eaglys currently is pursuing the hybrid develop and consult model that was discussed earlier herein.

Section 4. BUSINESS RISKS & MITIGATION

To analyze some of the risks that a company competing in the homomorphic encryption market faces we utilize a Cross SWOT instead of the more common SWOT approach to look into the various items that we have discussed throughout this paper and what kind of a threat profile they generate against the current capabilities and weaknesses of the business - in most cases the issues can be

mitigated with appropriate partnerships or preparation for change, however as in the case of all new entrants it is quire crucial to recognize that an attempt to safeguard against all possible threats at the same time is nearly impossible for new startup entrants as resources will be too limited, and too easily exhausted. Attached is the Cross SWOT matrix, due to its size it is only available down in the Appendix section as a single table, displayed diagonally.

To annotate the most outstanding points on performance, research points to a reasonable level of performance achievable by homomorphic encryption as per Louk and Lim's research in the research paper 'Homomorphic Encryption in Mobile Multi Cloud Computing'; and states that it is indeed sufficiently fast to encrypt data that is being processed in the cloud (Louk and Lim, 2015). This is further corroborated by Ishimaki, et. al. in the 2017 IEEE publication concerning 'Private Substring Search on Homomorphicaly Encrypted Data', where while it is presumed to be a reasonably resource intensive task as per the below:

> "In order to perform private substring search, Fully Homomorphic Encryption (FHE) can be adopted although it induces computationally huge overhead. Because of the huge overhead, performing private substring search efficiently over FHE is a challenging task." (Ishimaki, et. al. 2017)

It was proven to be optimizeable, and after analysis it was found that in essence the problem could be overcome:

> "[...] substring search conducted by FHE can be done within feasible time. We also ensure that the analyst does not know anything other than the searched result from the server due to our proposed Batch Recursive Oblivious Transfer. " (Ishimaki, et. al. 2017)

That said the above does point out that there is a concern in handling very large data sets in homomorphic encryption schemes, especially if results need to be retrieved in near-real time, as the database size could approach an increase in size by an order of magnitude - and that is noted in the Cross SWOT; the key to addressing this weakness is both partnering with cloud providers and potentially hardware device developers. This would allow the cloud provider to offer additional security as well as upsell reason for capacity and speed. On the other side under the hardware solution perspective, partnering with hardware design companies could provide both a cryptographic chipset design to be used on-site by various providers, as well as a cross sell opportunity for homomorphic encryption schemes from the hardware provider channel. This would of course require further research into optimization and hybridization of the approaches. Furthermore, as size and speed requirements increase, and if a cost ceiling is present, hybrid approaches can be considered to encrypt data in zones or levels depending on the criticality of exposure and risk, against the need to perform operations and the feature set required; a common approach to all limited resource problems.

The knowledge curve effect and size risk is another threat that becomes apparent in the Cross SWOT, as larger companies could potentially start developing their own solutions; however given the multitude of approaches and the relative early stage of the market this is an acceptable risk - additionally the large providers that would have the most to gain in integrating this technology into their core offering do have a tendency to acquire startups that offer unique technologies as we have found in the previous chapters; this can be considered to alleviate such risks to a degree, as well as a potential strength as a smaller company could adjust to demands quicker. This also offers a potential exit strategy for startups, as the technology developed in smaller firms may be desirable for inclusion in a larger firm's portfolio - especially if it enhances the offering on hand.

The last item that stands out is the regulatory exposure risk, as any privacy and encryption technology is subject to various legal forces; while the direct regulatory risk of the legality of specific

technologies cannot be directly addressed, some flexibility exists in the global market; this can also be offset by focusing on segments that are in dire need of the technology. On the other side of the equation, there does not seem to currently exist strict guidelines on which technology is to be used to secure data such as medical patient data, financial institution transaction data or the like. Given that the technology discussed herein is indeed proven to be secure, this should not pose an immediate threat in the near future - however it should as always be monitored closely.

As a side note, one risk that did not appear in the Cross SWOT directly as it could potentially fall under the regulatory or technology section, was the risk of the AI manipulation technology, whereas AI models can be manipulated to focus on specific results that the attacker wants the AI to identify instead of the correct answer. This is described as such in Ben Dickson's article 'Protecting AI models against audio adversarial attacks' as follows:

> "Adversarial examples make subtle changes to the input of a machine learning model in a way that causes its output to change. For instance, adding a layer of noise to the image of a panda will cause an AI to classify it as a gibbon. But to a human, it will still look the same." (Dickson 2019)

Given that the attacker could potentially make an AI 'believe' the wrong answer, attacks such as the above may cause some delay to deployment of AI in highly sensitive sectors - which would slow down market growth in the AI and machine learning segment as well as ancillary segments such as data security.

CHAPTER 5. CONCLUSION

Currently the data security market segment is quite vibrant and the startup community has taken notice; at the current moment and with the appropriate approach to avoid strategic mistakes a great opportunity exists to seize the medical data security market by using homomorphic encryption approaches to secure both the data storage and the data processing side of the equation. As to the potential down-the-road exit strategy for such a startup, two examples stand out that could be emulated; the Havok and GitHub acquisitions by Microsoft. The Havok game engine was acquired by Intel and then was sold onward to Microsoft in 2015, and notably when acquiring the company Microsoft's focus on announcing that they will continue developing those tools as they were developed by Havok:

> "Microsoft's acquisition of Havok continues our tradition of empowering developers by providing them with the tools to unleash their creativity to the world. We will continue to innovate for the benefit of development partners. Part of this innovation will include building the most complete cloud service, which we've just started to show through games like 'Crackdown 3.'
>
> Havok shares Microsoft's vision for empowering people to create worlds and experiences that have never been seen before, and we look forward to sharing more of this vision in the near future." (Microsoft, 2015)

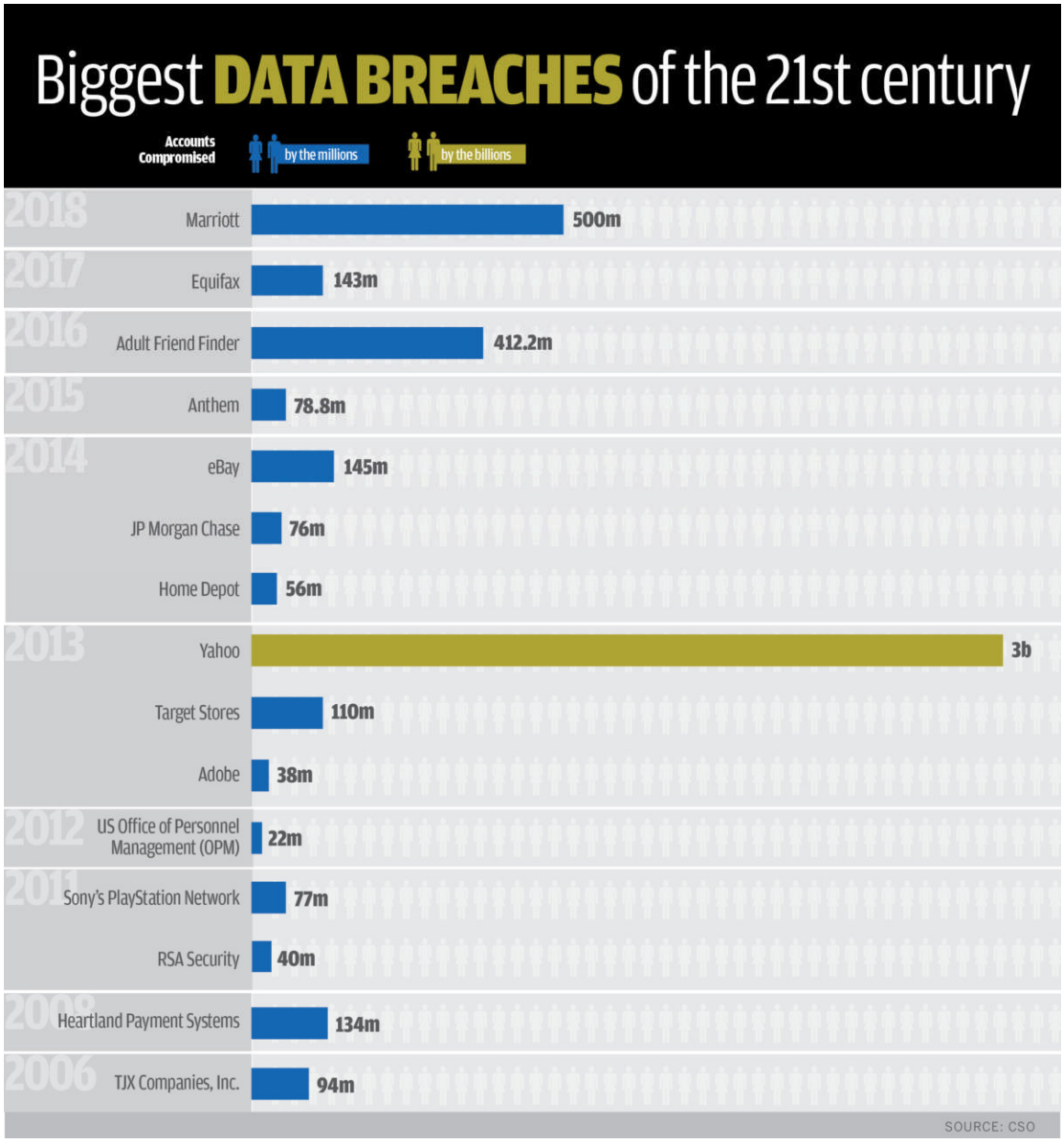
It appears that Microsoft's focus is to acquire and extend the tools so that they may offer them to their own customers, after integrating them into the Microsoft ecosystem.

The other company mentioned previously, GitHub, was acquired by Microsoft in 2018; similarly, Microsoft's public relations announcement focused on the continuation of the tool and working with the provider to further develop the product that they acquired in support of their own customers, as per the below mention:

> "GitHub will retain its developer-first ethos, operate independently, and remain an open platform. Together, the two companies will work together to empower developers to achieve more at every stage of the development lifecycle, accelerate enterprise use of GitHub, and bring Microsoft's developer tools and services to new audiences." (Microsoft, 2018).

In the long run, there is good potential for an acquisition exit for a startup in cases where the company develops a tightly integrated product and supports a large user base that overlaps with a large service provider's own user population; especially if the product offered enhances the customer experience, and the portfolio of the larger firm. Those in the field tend to take notice, and a partnership or an acquisition is definitely possible.

CHAPTER 6. APPENDIX



Recent Data Breaches, Source: Amerding, 2018

	Government Regulation	↑	and focus on open markets	↓	↑	with research organizations	↓
	Threat from large entities	Potential for acquisition	Focus on integration	custom solutions	computing partners	experience effect	Focus on innovation
Threats	Development bottleneck	R&D and PoC projects	ready-to-market first	consulting diffusion	↑	Scale up R&D	↓
	Local markets	advantage of local branding	local network power	added services	Partner with local startups	providers that lack expertise	Growth focus
	Hardware markets	hardware makers	(deploy and forget)	hardware makers	with hardware makers	engage high profile targets	low cost solutions
Opportunities	Cloud computing	could drive expansion	applicable to wide markets	of consulting stream	could drive sales	control market	to trends quickly
	SWOT	Data security	Integration (DB+AI)	Customization	Computing requirements	Relatively new market	Small scale development
CROSS		Strengths			Weaknesses		

Cross SWOT, Source: Author, 2019

REFERENCES

[1] Amazon (2019) "About AWS", AWS, Amazon, https://aws.amazon.com/about-aws/

[2] Amazon (2019) "Amazon.com Announces Fourth Quarter Sales up 20% to $72.4 Billion", The Amazon Blog, Amazon, January 2019, "https://ir.aboutamazon.com/news-releases/news-release-details/amazoncom-announces-fourth-quarter-sales-20-724-billion

[3] AWS (2012) "Amazon DynamoDB Developer Guide (API Version 2012-08-10)", Amazon, August 2012 https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/EncryptionAtRest.html

[4] Bhattacharya, S. (2019) "The New Dawn of AI: Federated", Towards Data Science, January 2019, Learninghttps://towardsdatascience.com/the-new-dawn-of-ai-federated-learning-8ccd9ed7fc3a

[5] Business World (2019)"Global AI-As-A-Service Market to Reach USD 14.71 Billion by 2024." Business World 18 July 2018. Business Insights: Global. Web. 1 July 2019, http://bi.galegroup.com/global/article/GALE%7CA546859640

[6] Callas, J., Donnerhacke, L., Finney, H., and R. Thayer, (1998). "OpenPGP Message Format", RFC 2440, https://doi.org/10.17487/RFC2440, November 1998

[7] Canalys (2019) "Canalys: battle for enterprise cloud customers intensifies as spending grows 42% in Q1 2019" https://www.canalys.com/newsroom/canalys-battle-for-enterprise-cloud-customers-intensifies-as-spending-grows-42-in-q1-2019

[8] CloudEndure "Products", CloudEndure, 2019 https://www.cloudendure.com/

[9] Dellinger, A.J. (2019) "Quest Diagnostics Hit with Class Action Lawsuit Over Breach That Exposed Patient Data", Forbes, June 2019, https://www.forbes.com/sites/ajdellinger/2019/06/12/quest-dynamics-hit-with-class-action-lawsuit-over-breach-that-exposed-patient-data/#389c00fb37b9

[10] Dickson B. (2019) "Protecting AI models against audio adversarial attacks", TechTalks, April 2019, https://bdtechtalks.com/2019/04/29/ai-audio-adversarial-examples/

[11] Dierks, T. and E. Rescorla (2008). "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, https://doi.org/10.17487/RFC5246, August 2008,

[12] Doel T. Shakir D.I. Pratt R. Aertsen M. Moggridge J. Bellon E. David A. L. Deprest J. Vercauteren T. Ourselin S. (2016). "GIFT-Cloud: A data sharing and collaboration platform for medical imaging research", Computer Methods and Programs in Biomedicine Volume 139, February 2017, Pages 181-190. https://doi.org/10.1016/j.cmpb.2016.11.004

[13] European Commission (2017) " Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679", EC Justice and Consumers, October 2017, https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

[14] Farrell, S. Tschofenig, H. (2014) "Pervasive Monitoring Is an Attack", BCP 188, RFC 7258, https://doi.org/10.17487/RFC7258, May 2014,

[15] Giarratana M.S. (2004). "The birth of a new industry: entry by start-ups and the drivers of firm growth - The case of encryption software", Research Policy 33 (2004) 787–806. http://doi.org/10.1016/j.respol.2004.01.001

[16] Google Cloud (2017) "Encryption at Rest in Google Cloud Platform - Google Cloud Platform Encryption Whitepaper", April 2017 https://cloud.google.com/security/encryption-at-rest/default-encryption/ https://cloud.google.com/security/encryption-at-rest/default-encryption/resources/encryption-whitepaper.pdf

[17] Hauben, M. (1997). "Netizens: On the History and Impact of Usenet and the Internet", May 1997, Wiley-IEEE Computer Society Press, published by author http://www.columbia.edu/~hauben/Book_Anniversary/

[18] Ishimaki, Y. Imabayashi, H. Yamana, H. (2017) Private Substring Search on Homomorphically Encrypted Data, IEEE International Conference on Smart Computing, 2017. SMARTCOMP 2017, IEEE, 29-31 May 2017, https://doi.org/10.1109/SMARTCOMP.2017.7947038

[19] Janakiram M. S. V. (2018) "The Rise of Artificial Intelligence as A Service in The Public Cloud", February 2018, Forbes https://www.forbes.com/sites/janakirammsv/2018/02/22/the-rise-of-artificial-intelligence-as-a-service-in-the-public-cloud/#3e17a085198e

[20] Louk, M. Lim, H. (2015). "Homomorphic Encryption in Mobile Multi Cloud Computing", 2015 International Conference on Information Networking (ICOIN) IEEE, 12-14 Jan. 2015. http://doi.org/10.1109/icoin.2015.7057954

[21] M2 Presswire (2019) "Global Cloud Computing Market Outlook to 2023: The Increase in Adoption of Hybrid Cloud Services Presents Lucrative Opportunities." M2 Presswire 23 May 2019. Business Insights: Global. Web. 1 July 2019, http://bi.galegroup.com/global/article/GALE|A586360508/ee1449de4aa0aa5483c408395670f1ec

[22] Martins, P Mariano, A (2017). "A Survey on Fully Homomorphic Encryption: An Engineering Perspective", ACM Computing Surveys, Vol. 50, No. 6, Article 83. Publication date: December 2017. https://doi.org/10.1145/3124441

[23] Mascha, M.F. Miller, C.L. & Janvrin, D.J. (2011). "The effect of encryption on Internet purchase intent in multiple vendor and product risk settings", Electron Commer Res 11: 401. https://doi.org/10.1007/s10660-011-9080-6

[24] Masinter, L., "Hyper Text Coffee Pot Control Protocol (HTCPCP/1.0)", RFC 2324, https://doi.org/10.17487/RFC2324, April 1, 1998

[25] Masood, I. Wang, Y. Daud, A. Aljohani, N. R. and Dawood, H. (2018). “Towards Smart Healthcare: Patient Data Privacy and Security in Sensor-Cloud Infrastructure", Wireless Communications and Mobile Computing, vol. 2018, Article ID 2143897, 23 pages, 2018. https://doi.org/10.1155/2018/2143897

[26] Matthews, L. (2016) "Adult Websites Hacked, 412 Million Users Exposed", Forbes, November 2016, https://www.forbes.com/sites/leemathews/2016/11/13/412-million-users-exposed-as-adult-websites-hacked/#35abe9e46c76

[27] Microsoft Corporate Blogs (2015) "Havok to join Microsoft", Microsoft, Microsoft Corporate Blogs, October 2015, https://blogs.microsoft.com/blog/2015/10/02/havok-to-join-microsoft/

[28] Microsoft Corporate Blogs (2018) "Microsoft completes GitHub acquisition", Microsoft, Microsoft Corporate Blogs, October 2018, https://blogs.microsoft.com/blog/2018/10/26/microsoft-completes-github-acquisition/

[29] O'Flaherty K. (2019) "Marriott CEO Reveals New Details About Mega Breach", Forbes, March 2019, https://www.forbes.com/sites/kateoflahertyuk/2019/03/11/marriott-ceo-reveals-new-details-about-mega-breach/#1f0a46bc155c

[30] Ojala, A. (2016). "Adjusting software revenue and pricing strategies in the era of cloud computing". The Journal of Systems and Software. https://doi.org/10.1016/j.jss.2016.08.070

[31] Owler (2019) "Competitors, Revenue, Number of Employees, Funding and Acquisitions", Owler, July 2019, https://www.owler.com/

[32] Postel, J. (1981). "Transmission Control Protocol", STD 7, RFC 793, https://doi.org/10.17487/RFC0793, September 1981,

[33] Roy, M. Mali, K Chatterjee, S. Chakraborty, S. Debnath, R. Sen, S. (2019). "A Study on the Applications of the Biomedical Image Encryption Methods for Secured Computer Aided Diagnostics", 2019 Amity International Conference on Artificial Intelligence (AICAI) IEEE, 4-6 Feb. 2019. https://doi.org/10.1109/AICAI.2019.8701382

[34] Stalcup, K (2019) " AWS vs Azure vs Google Cloud Market Share 2019: What the Latest Data Shows", Park My Cloud, April 2019, https://www.parkmycloud.com/blog/aws-vs-azure-vs-google-cloud-market-share/

[35] Taylor Armerding (2018) " The 18 biggest data breaches of the 21st century", CSO, December 2018, https://www.csoonline.com/article/2130877/

[36] Technavio (2019) "Healthcare Cloud Computing Market Is Witnessed to Grow USD 16.37 Billion, at 22% CAGR from 2018-2022" /. Business Wire 28 June 2019. Business Insights: Global. Web. 1 July 2019, http://bi.galegroup.com/global/article/GALE|A591207227/ee1449de4aa0aa5483c408395670f1ec

[37] Tibouchi, M. (2014) "Fully Homomorphic Encryption over the Integers: From Theory to Practice", July 2014, NTT Technical Review Vol. 12 No. 7 https://www.ntt-review.jp/archive/

[38] TSO Logic, "About Us", TSO Logic, 2019 https://tsologic.com/about-us/

[39] U.S. Department of Health and Human Services (2013), "HIPAA Administrative Simplification" U.S. Department of Health and Human Services Office for Civil Rights March 2013, https://www.hhs.gov/sites/default/files/hipaa-simplification-201303.pdf

[40] Weaver, A.C. Dwyer, S.J. Snyder, A.M. Van Dyke, J. Hu, J. Chen, X. Mulholland, T. Marshall, A. (2003) " Federated, secure trust networks for distributed healthcare IT services", IEEE International Conference on Industrial Informatics, 2003. INDIN 2003. Proceedings, IEEE, 21-24 Aug. 2003, https://doi.org/10.1109/INDIN.2003.1300264

[41] Wolverton, T. Lee, S. "The cloud-computing market is set to double to $116 billion by 2021 — these 3 charts show why that's probably good news only for Amazon, Google, Microsoft, and Alibaba" Business Insider, November 2018 https://www.businessinsider.com/goldman-sachs-cloud-computing-market-forecast-aws-microsoft-azure-google-cloud-2018-11

[42] Yuan, J. Malin, B. Modave, F. Guo, Y. Hogan, W.R., Shenkman E., Bian, J. (2017) " Towards a privacy preserving cohort discovery framework for clinical research networks" Journal of Biomedical Informatics Volume 66, Pages 42-51, February 2017, https://doi.org/10.1016/j.jbi.2016.12.008

[43] ZDNet (2019) "These are the worst hacks, cyberattacks and data breaches of 2019 so far", ZDNet June 2019, https://www.zdnet.com/pictures/these-are-the-worst-hacks-cyberattacks-and-data-breaches-of-2019-so-far/