Achieving GDPR Compliance with Unikernels

Bob Duncan
Computing Science
University of Aberdeen, UK
Aberdeen, UK
Email: bobduncan@abdn.ac.uk

Andreas Happe
Dept. Digital Safety & Security
Austrian Inst. of Tech. GmbH
Vienna, Austria
Email: andreas.happe@ait.ac.at

Alfred Bratterud
Dept. of Computer Science
Oslo and Akershus University
Oslo, Norway
Email: alfred.bratterud@hioa.no

***Abstract*—IT security and privacy has always been a challenging problem to address, but with cloud, there is an exponential increase to the challenge. Once an attacker successfully breaches a cloud system, the intruder will seek to escalate privileges in order to delete the forensic trail, thus covering their tracks. There is little to prevent this from happening in cloud, and this is known as the Cloud Forensic Problem. Under the new European Union General Data Protection Regulation, following a cyber breach, it is necessary for the breached company to report the impact of the breach within 72 hours of becoming aware of the breach. Where the forensic trail has been compromised, this will present a serious compliance challenge. We address this problem through the use of Unikernel based monitoring systems which can ensure both full forensic and audit trails can be maintained. Our early results are very promising. We are continuing our work with a larger pilot study.**

Keywords–Cloud Forensic Problem; unikernels; EU GDPR, compliance.

I. INTRODUCTION

All business is the subject of cyber attacks, no matter whether it is a public corporation, a private firm, a financial institution, a government agency, a non-government agency or a charity. In previous work [1], we proposed the use of a unikernel based system to help defend against such attacks. No matter what type of organisation is involved, all those who will be subject to the rules of the European Union (EU) General Data Protection Regulation (GDPR) [2], will need to comply fully with the regulation. No matter where the company is located in the world, should they hold personally identifiable information (PII) belonging to any EU resident, they will fall under the jurisdiction of the EU GDPR regulator. In a post-Brexit world, the UK Government has indicated that the GDPR will still apply in the UK. Indeed, the UK Government has indicated that the UK GDPR will be enforced with greater rigour, and will accord greater rights to private individuals.

In order to achieve compliance with the rules of the GDPR, companies who fall under the scope of the GDPR will necessarily require to undertake considerable extra work, and expense, in order to be able to achieve compliance. Each organisation will require to appoint a data controller, who either must have the requisite technical skills, or must be assisted by a person with such technical skills. This will likely be an unwelcome additional expense. They must also have a data processor and a data protection officer, meaning further costs. In addition, they will have to take all necessary technical steps to ensure the security and privacy of all PII belonging to data subjects of the organisation, again at yet more expense.

Many companies are likely to be unprepared for achieving compliance. Many (erroneously) believe that because the reporting requirement has been changed from "within 72 hours of a breach occurring" to "within 72 hours of discovering a breach", they will have no problem being compliant [3]. The reality is that they will be wrong! They must also be able to report which records were accessed, modified, deleted or exfiltrated from the system. However, once an attacker breaches a system and becomes resident as an intruder, one of the first tasks they seek to carry out is to delete the forensic trail which recorded their incursion into the enterprise systems, so that their presence becomes more covert, allowing them to remain hidden inside the system. This allows them to harvest whatever information or secrets they desire for as long as they remain hidden in the system.

Without a complete forensic trail in any system, compliance will be a challenge, if not impossible. This will particularly be the case with cloud systems, since there is nothing to prevent such an intruder from deleting not only the forensic trail, but anything else they desire, including the very running cloud instance that they are hiding within. If there is no record of the trail of events relating to the database contents, then the company is unlikely to be able to identify which records have been accessed, modified, or deleted, resulting in a failure to be compliant with the GDPR. Since failure to comply can result in fines which can rise to the greater of €20 million or 4% of global turnover, then this will certainly have a substantial impact, although there are many who still fail to grasp this important point.

We start by considering the cloud forensic problem in Section II, and discuss why this is such a challenge for GDPR compliance in cloud systems. We are concerned with achieving both good security and good privacy. While it is possible to have security without privacy, it is not possible to have privacy without security. Thus our approach will be to first ensure a good level of security can be achieved, and to that end, we start by listing the specific security issues we seek to address and discuss how we propose to tackle them in Section III. The remainder of the paper is organized as follows: in Section IV, we consider how we might go about finding a cloud based solution, in Section V, we discuss the outline technical details of our proposed approach; In Section VI, we consider possible attack vectors. In Section VII, we consider just how robust a unikernel approach might be. In Section IX, we discuss our conclusions.

II. THE CLOUD FORENSIC PROBLEM AND THE GDPR

Cloud computing has been around now for over 10 years, and a great deal of good quality research has been carried out,

particularly regarding matters of security and privacy. Cloud systems have become highly popular with companies due to the flexibility of cloud offerings. The speed of starting a cloud instance, the removal of long lead times to access compute and storage resources, the ability to scale up, as well as down, to match demand presents a particularly good incentive to use cloud computing. The fact that companies can write costs off entirely against revenue provides a further attractive incentive for their use. Kratzke [4] has long warned of the dangers of thinking that conventional software is just the same as cloud-native software. Kratzke et al. [5] do suggest the possibility of using existing Container technologies to improve cloud-native programming.

There have been many good papers produced on both security [6]–[17] and privacy [14], [18]–[32], and we laud the efforts of countless researchers who have tried to provide this area with better security and privacy, which speaking generally, has been successfully achieved over the years. A number of others have looked at better accountability as a means to meeting these ends [10], [11], [15], [20], [27], [30], [33]–[52] But there remains one fundamental weakness that has not been resolved, namely the "cloud forensic problem". All computer systems are the subject of attack, and cloud systems are no exception. Unfortunately, no system is immune to attack, and that is certainly true for cloud systems.

No computer system is immune to attack. It is the primary goal of an attacker to breach a system. This can involve quite a considerable amount of work on the part of a serious attacker. They are very likely to perform extensive surveillance and compile many analyses of how the company systems and their architectures are organised. Many will carry out considerable amounts of social engineering work to attempt to fully understand the people of the organisation, since people are frequently the weakest link. But understanding the organisational structure can also provide vital intelligence to understand how company procedures operate, all of which can help them achieve their goals. This intelligence gathering will be very comprehensive and thorough, usually covering every possible aspect of all the systems of the company in order to discover everything they can about the business architecture before they start their attacks. By understanding fully how the company is structured and how it operates, they are far less likely to make any errors when they start the process of penetrating the systems.

Other attackers, are much less organised. They will simply try to hack in to company systems, without any regard or thought of the overview of the company concerned. They will merely look for known software vulnerabilities and try their best to successfully attack them. They care little about whether they are discovered while attempting to penetrate the system. Theirs is a short term view, rather than the long view held by others. They want to get in, and out, quickly with whatever they can lay their hands on. For them, time is money, and if they are unable to get in within a reasonable amount of time, they will move on to the next prospect.

Yet other potential intruders will perpetrate their attack through the people of the business using a variety of other attack methods: such as using social engineering attacks, email attacks that might use malicious links and malware payloads, attempting to use web based drive by attacks, or the use of phishing, vishing and many other approaches. These attackers are much less concerned with purely technical attacks, but are often extremely talented in the use of these methods, and in particular social engineering.

No matter which type of attacker they are, they all share one fundamental goal — and that is to penetrate the system in order to become an intruder. The aim here is not just to get in, and out, as quickly as possible, but to develop a long term foothold inside company systems which will allow them to return, time and again, to help themselves to whatever they wish, as they escalate privileges more and more, the longer they remain inside the system. This will necessarily involve some serious attempts to escalate privileges to allow them to mdify the forensic trail.

It is rather unfortunate that they are often greatly aided in this quest by the companies themselves. Usually, this occurs through a degree of laziness whereby the companies are clearly failing to monitor server logs properly. Looking at previous cyber breach reports [53], at which time there was a global average of 6 months between breach and discovery, it is clear that had these companies been more rigorous about reading and analysing their server logs, they would have been in a better position to discover intruders rather more quickly. Even last year, where the time between breach and discovery has dropped to a number of weeks rather than months [54], this is still not good enough. Some companies contribute further by refusing or failing to update security patches to both operating systems and software systems, usually under the guise of "last time I did a security update, all the systems crashed".

This all leads towards the, as yet unresolved, cloud forensic problem — namely, that once an intruder is in the system, and has escalated sufficient privileges, there is absolutely nothing to prevent them from deleting the forensic trail, which allows them to hide all evidence of their successful penetration. Worse, by this stage they will also have control of all the system logs and audit trails, and there is nothing to prevent them from deleting every last trace of their intrusion and ongoing ex-filtration of private data.

Surely that has nothing to do with the GDPR you might ask? Sadly, that is not the case. In the event of a breach, you are required to report the breach within 72 hours of discovering the breach. You must be able to report how many relevant records have been compromised, whether by having been read, amended, deleted or ex-filtrated from the system. Given that many system logs are not even turned on by default, this means identifying which records have been compromised, whether by having been read, amended, deleted or ex-filtrated from the system, will present a serious enough challenge in the first place. However, given that the intruder will likely have thoroughly worked through all forensic trails in the system, the likelihood of being able to realise that a breach has occurred at all will likely be very slim, let alone having the ability to properly identify which records have been compromised.

From a holistic perspective, it would have been helpful if these matters might have been addressed by the Cloud implementation itself. However, no such attempt has taken place during the past decade, no doubt due to the massive challenge involved. Consequently, all organizations subject to the provisions of the GDPR are required to safeguard their own systems and therefore take such steps as are necessary to ensure adequate privacy is achieved.

This will mean non-compliance with the GDPR, which can then trigger fines which can rise to the greater of €20,000,000 or 4% of global turnover. This will certainly catch the attention of top management within organisations. Considering that these fines can be levied for every single breach, and that the upper limit is based on turnover rather than profit, that should be sufficiently concerning to get some serious attention. Of course, all sensible Cloud users should have been thinking about this long before now, and we are aware of many who on hearing that notification 'within 72 hours of discovery of a breach', rather than 'within 72 hours of occurrence of a breach', heaved a collective sigh of relief and stopped worrying about implementing a solution. This is what motivates our work.

III. How do we tackle the problem?

At this time, no system is fully secure. Operating systems, transport protocols, software applications — all of this software has evolved during previous decades. Any relevant standards were defined decades ago. The primary goal at that time was functionality. Security and privacy were very much an afterthought, which has remained the case for decades. Security and privacy has very much been a case of "Let us bolt something on to tackle that". Default settings are geared for ease of setting up, not for security and privacy. This means proper security and privacy presents a massive challenge, which increases exponentially for cloud, Internet of Things and Big Data.

Since the primary goal of the successful intruder is to delete or obfuscate the forensic trail which could expose their presence, then we must consider protection of this data a priority. However, before becoming a successful intruder, the attacker has first to get into the system. This process will be capable of triggering certain alarms, if activated. At the very least, proper scrutiny of server logs would be a big help here. It is not necessary to have human eyes on all these logs, but it would be sensible to use some automated means to detect anomalous behaviour and to flag this up before the attacker can gain a permanent foothold within the system. Thus, there are two specific needs to fulfil here. One is the proper protection of all forensic data and audit trails, and the other is to analyse the system traffic in a timely manner to detect potential anomalous behaviour.

One might imagine that it would logically be more efficient to deal with the second need first before considering the first. However, as we have already stated, retaining a full and proper record is not only vital for GDPR compliance requirements, but with compromised forensic and audit trails, there will not be a full picture to analyse for anomalous behaviours, rendering the task less than useful. We therefore suggest the protection of the forensic and audit data has to be the priority, meaning that the subsequent analysis of this data will at least be run on a full set of data.

We therefore address the security of the forensic and audit trail data as our first priority, returning to the analysis of log data to detect anomalous behaviour in Section VIII. We therefore seek a suitable mechanism that will be fit for our purposes, and consider here the advantages and disadvantages of a number of possible alternatives.

Conventional algorithms running on the server could potentially work well, but their weakness lies in running on the server instance where they are vulnerable to attack. They would also present a considerable overhead to the smooth running of the main web application on the cloud instance.

We could opt to use Containers, such as Docker, LXD or Rocket. However, Bratterud et al. [55] warn of some security issues with this approach, and Kratzke [56] also warns of the unexpected, and unwelcome overhead these solutions can bring.

In previous work, [57], we considered how well unikernels might be used to improve on dealing with our target list of security goals, and found the potential for an improved approach. In [58], we developed a suitable framework, providing detailed definitions of how this might be tackled. In [59], we demonstrated how a unikernel based solution could reduce complexity, while improving security and privacy. We also considered in [60], whether unikernels could help address some of the key weaknesses introduced by use of the Internet of Things (IoT). In each case, we build on the work of the previous papers, in order to ensure we do not miss anything important as we develop the system.

Unikernels run natively on cloud, they have an exceptionally small footprint, they run without many of the conventional "toys" associated with normal web based cloud instances. This means a seriously minimal attack surface. They are lightweight, can be activated "on demand", and are extremely difficult to attack. Virtually every single conventional attack fails due to there being a heavily restricted means of accessing the running unikernels. A typical cloud instance will be over 150MB in size. Even Docker containers will be a minimum of 24MB in size, whereas a unikernel instance can be as little as 2MB in size. This approach is therefore of interest to us in working towards a good solution to the problem.

Given the limitation we face in terms of most software being insecure, how can we approach developing a potential solution for this problem? In [61] [62] Duncan and Whittington proposed that all cloud based systems which would be subject to compliance under the GDPR, should have ALL audit trails and forensic logs captured and stored off-site in a highly secure immutable database running on a dedicated and highly secure server. These proposals also suggested the immutable database be set up off-site from the cloud instance. This solution has the advantage that the data is not available on the running cloud instance for an attacker to try to compromise, leading to a more secure approach.

While we accept that advice might be highly appropriate given the pervasive extent of the cloud forensic problem, could there be any other way that we might be able to find a cloud based solution? As we shall see in the next section, there may be a way to achieve just that objective.

IV. Finding a Cloud Based Solution

We certainly do accept the sensible logic proposed by Duncan and Whittington [61] [62] to keep the immutable database separated from all running cloud instances. While that makes perfect sense, there is no reason, other than the cloud forensic problem, why the immutable database should not run on a cloud system. However, we do agree that it should not run on the same system as the company system it is trying to protect.

We are keen to explore the idea of running a system on cloud, since that will have the attraction of having all the characteristics that make cloud an interesting proposition for enterprises to use. It provides an agile way to match demand needs to the supply of resources, which can be acquired on demand. It is highly flexible and infinitely scalable. When provisioned by a serious CSP, it is likely to be much more secure than a conventional distributed network system that has been poorly configured. It is also a revenue expense, which can be advantageous for fiscal reasons.

So the question we must now address is how we might go about solving this particular problem. This is where the unikernel based system might be able to help.

Let us first consider the advantages from a security point of view of unikernels:

- The larger a piece of software, the more vulnerabilities are usually present. As we already stated, a unikernel instance can be as little as 2MB;
- The smaller an instance is, the faster a new instance loads;
- The smaller instances are, the cheaper they are to run;
- There is no terminal to log into. The terminal presents one of the easiest attack routes into any system and is usually not well protected from attack. If the attacker cannot log in, achieving a successful attack will be rather difficult to perpetrate;
- The running instance of any unikernel runs with immutable code, meaning no user may inject code into the running unikernel instance.

And now, let us look at any potential disadvantages of unikernels:

- No terminal to log into — a disadvantage for most sys admins. We view this as a huge advantage. If the sys admin cannot login, the attacker has no chance of doing so;
- The running instance is immutable, so it cannot handle changes. We view this as a positive. We are particularly keen to be able to log all changes, system, forensic and audit trail data in a persistent and immutable storage medium off-site. If we cannot change anything, neither can the intruder.

In our view, every item in the above list of advantages and disadvantages all present positive attributes. From a performance, cost, reduced latency and minimised attack surface perspective, all the attributes are highly beneficial for our purposes. This provides us with a degree of confidence that we might be moving on the right track to find a workable solution.

In the next section, we will look at how we might set about developing a system to deploy these instances in a suitable manner that might help us to solve our security challenge.

V. Outline Technical Solution Proposed

We have seen that our unikernel instances can be extremely lightweight, are immutable in operation, have a very small attack surface, perform well, are cheap to run with reduced latency. Because of these advantages, we can use a number of these instances to build a much more robust system.

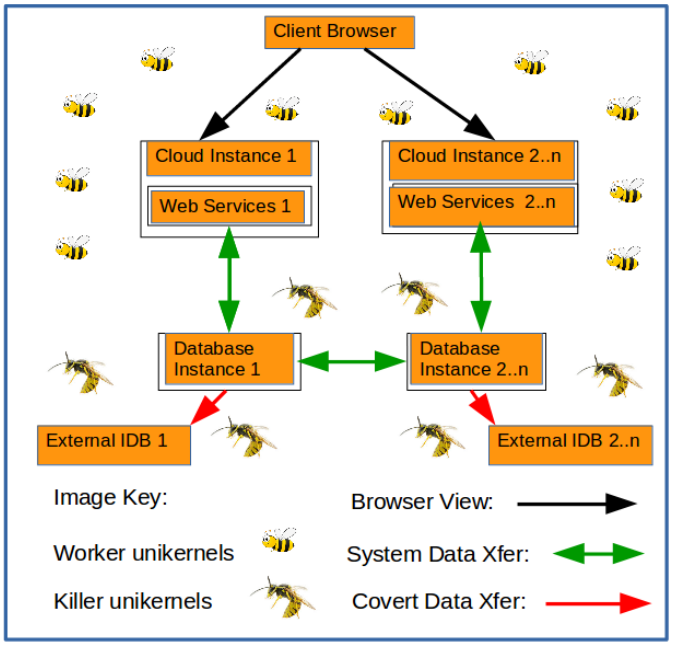

Figure 1: A Unikernel Based Solution to the Cloud Forensic Problem.

If we use the analogy of a bee hive, we can apply this approach as part of our solution. In a bee hive, there are a number of specialised bees — there is a single queen bee, hundreds of male drones (whose responsibility is to mate with a queen, after which they die), anything up to 80,000 female worker bees, who look after developing eggs, larvae and pupae, as well as the whole hive, gathering food from flowers outside the hive and defence duties, which they perform to the death, if needed. Each bee performs a specialised function depending on its age. And in the event a queen leaves, gets lost, or dies accidentally, the colony is capable of generating new queens, either full queens, or temporary queens. The ultimate in sustainability.

Our main company system will have a presence on a cloud platform, using one or more cloud instances as needed, which will be running on a conventional cloud setup. The cloud instance will have the capability to replicate at scale as demand increases and also to shut down instances when demand falls. The main cloud instance system will not be able to be shut down from within. We shall call this the front end Cloud Instance 1.

A conventional database management system will be included in all cloud instances in the normal way except they will instead be removed from within these instances and will run inside a single instance with every non-required function removed from that running instance in order to reduce the attack surface. Should database replication be later required, this can be accommodated through setting up similar database instances. We shall call this original Database Instance 1.

Thus Database Instance 1 will only accept input from the known running front end Cloud Instance 1. There will be no direct access allowed from outside the cloud environment. In the event that replication is required, Cloud Instance 1 will setup as many replicated instances as needed, including

Database Instance 2..n, which will all be replicated, expanding to deliver the required resources.

Worker unikernels will be assigned to each Cloud Instance as they are spooled up, and shut down as no longer needed. They will have specific tasks to perform, such as policing, audit, or whatever. Killer unikernels will be assigned to the task of protecting database systems. Their primary goal will be to ensure the safety of both the forensic trail and the audit trail for all database components, which will be safely stored in the immutable database. These records cannot be deleted. If required, these killer unikernels can turn on attackers trying to breach the systems. All unikernel instances will be tracked, with forensic data collected also for them.

As we can see, each different type of instance is specialised, sticking to its own designated tasks. So what is special about this, apart from splitting up the functions? When a cloud instance runs with a variety of different types of software running on it, this can present a big challenge to configure the overall package in a secure way. By specialising each instance, it becomes much easier to configure securely, because every single unused port can be shut down. Security controls can focus on only what they have to, thus cutting down the potential attack surface.

Any new front end instance, if not registered with the control instance, will not be allowed access to any database instance. Likewise where any new database instance is not registered with the control instance, the front end instances will refuse to connect with it.

The secure immutable database for storing system logs, forensic and audit trail data should not be directly visible to the client browser. Each running instance will send a copy of all system logs, forensic and audit trail data to the immutable database instance as it is generated. The source and timing of all events will be logged by the immutable database.

With the unikernel instances, because they are so lightweight, we can deploy as many of them as we need to carry out very specific tasks. We can have some to police various events, some to carry out audit tasks, some to keep a track of what is live within the system. Each of the components of the main system can be looked after by a number of dedicated unikernel instances, whose sole function will be dedicated to looking after the one conventional cloud instance. Since these unikernels are self sufficient, there is unlikely to be any real adverse impact on the running main instances.

Figure 1 shows a cross-section of the proposed solution. The Client browser will see the front end which provides conventional running cloud instances, with controllers hidden behind the scenes. These controllers can be protected by 'killer bee' unikernels. The external Immutable Database instances will be hosted elsewhere, and can also be protected by 'killer bee' unikernels. The 'worker bee' unikernels clustering around the conventional cloud instances will carry out normal policing and other required tasks. Additional 'bee workers' of whatever kind needed can be spooled up as required. They are fast to provision, take little space and will carry out their programmed task.

As to the question of how many of each type of unikernel we should aim to use, we believe that it would be pointless to speculate at this stage until we can test what will be optimal after we carry out some live experimentation to establish what works well in various loading scenarios. With the use of proper control systems, we can ensure that each new instance is properly registered, constantly and properly monitored, with the control system having the capability to spool up new instances as needed qhickly and efficiently, as well as shutting down those which are no longer required. We expect that such flexibility will allow a highly scalable system to be developed, which can offer minimal running cost, in conjunction with a low latency approach to dealing with attacks. This testing will form part of our future work.

VI. Possible Attack Vectors to Consider

Since we are mostly working with web services, we will look at the Open Web Application Security Project (OWASP) 2017 Top 10 Web Vulnerabilities [63]. We choose these, because they represent the top 10 vulnerabilities with the biggest financial impact on web user systems.

A1:2017-Injection Vulnerability: Injection flaws, such as Structured Query Language (SQL), Not Only SQL (NoSQL), Operating System (OS) injection and Lightweight Directory Access Protocol (LDAP) injection, occur when untrusted data is sent to an interpreter as part of a command or query. The attacker's hostile data can trick the interpreter into executing unintended commands or accessing data without proper authorization. **Solution:** Use a strong Application Programming Interface (API), separate content from commands in the database, and sanitise **ALL** user input.

A2:2017-Broken Authentication Vulnerability: Application functions related to authentication and session management are often implemented incorrectly, allowing attackers to compromise passwords, keys, or session tokens, or to exploit other implementation flaws to assume other users' identities temporarily or permanently. **Solution** Implement multi-factor authentication; no default passwords, especially from admins; reject all top 10,000 worst passwords.

A3:2017-Sensitive Data Exposure Vulnerability: Many web applications and APIs do not properly protect sensitive data, such as financial, healthcare, and PII. Attackers may steal or modify such weakly protected data to conduct credit card fraud, identity theft, or other crimes. Sensitive data may be compromised without extra protection, such as encryption at rest or in transit, and requires special precautions when exchanged with the browser. **Solution:** Encrypt all PII.

A4:2017-XML External Entities (XXE) Vulnerability: Many older or poorly configured eXtensible Markup Language (XML) processors evaluate external entity references within XML documents. External entities can be used to disclose internal files using the file Uniform Resource Identifier (URI) handler, internal file shares, internal port scanning, remote code execution, and denial of service attacks. **Solution:** Whenever possible, use less complex data formats such as JavaScript Object Notation (JSON), and avoiding serialization of sensitive data.

A5:2017-Broken Access Control Vulnerability: Restrictions on what authenticated users are allowed to do are often not properly enforced. Attackers can exploit these flaws to access unauthorized functionality and/or data, such as access other users' accounts, view sensitive files, modify other users' data, change access rights, etc. **Solution:** With the exception of public resources, deny by default; no unrestricted access to users; log all failures.

A6:2017-Security Misconfiguration Vulnerability: Security misconfiguration is the most commonly seen issue. This is commonly a result of insecure default configurations, incomplete or ad hoc configurations, open cloud storage, misconfigured Hypertext Transfer Protocol (HTTP) headers, and verbose error messages containing sensitive information. Not only must all operating systems, frameworks, libraries, and applications be securely configured, but they must be patched/upgraded in a timely fashion. **Solution:** Secure installation processes should be implemented. Keep it simple and log all errors.

A7:2017-Cross-Site Scripting (XSS) Vulnerability: XSS flaws occur whenever an application includes untrusted data in a new web page without proper validation or escaping, or updates an existing web page with user-supplied data using a browser API that can create Hyper Text Markup Language (HTML) or JavaScript. XSS allows attackers to execute scripts in the victim's browser which can hijack user sessions, deface web sites, or redirect the user to malicious sites. **Solution:** Preventing XSS requires separation of untrusted data from active browser content.

A8:2017-Insecure Deserialization Vulnerability: Insecure de-serialization often leads to remote code execution. Even if de-serialization flaws do not result in remote code execution, they can be used to perform attacks, including replay attacks, injection attacks, and privilege escalation attacks. **Solution:** The only safe architectural pattern is not to accept serialized objects from untrusted sources or to use serialization mediums that only permit primitive data types.

A9:2017-Using Components with Known Vulnerabilities Vulnerability: Components, such as libraries, frameworks, and other software modules, run with the same privileges as the application. If a vulnerable component is exploited, such an attack can facilitate serious data loss or server takeover. Applications and APIs using components with known vulnerabilities may undermine application defenses and enable various attacks and impacts. **Solution:** There should be a patch management process in place to ensure known vulnerabilities are never used.

A10:2017-Insufficient Logging & Monitoring Vulnerability: Insufficient logging and monitoring, coupled with missing or ineffective integration with incident response, allows attackers to further attack systems, maintain persistence, pivot to more systems, and tamper, extract, or destroy data. Most breach studies show time to detect a breach is over 200 days, typically detected by external parties rather than internal processes or monitoring. **Solution:** This paper is all about solving this problem!

And for no 11 of 10, go check out your site and make sure your system is not vulnerable.

There are, of course, many more vulnerabilities you can check out, and you should. The more you eliminate, the stronger and more robust your system becomes. You can be sure the attacker already knows all the potential vulnerabilities, so you need to make sure you do too, and plug them.

VII. Discussion on Outline Technical Solution

We strongly believe that a unikernel based system would have a positive and robust impact because of the extra muscle offered to check and log everything that is happening within the system. Given that unikernel instances have a very low attack surface, no conventional attacker 'toys', are immutable in operation, and highly compact, as well as everything being logged to the immutable database - we are cutting out a huge range of vulnerabilities from existing cloud systems. By ensuring the cloud instance running can also withstand the OWASP top ten web vulnerability test, we are in a very strong position to resist a great many attacks.

Some experimentation will be required to identify what the optimal setup of the 'unikernel hive' instances will be in order to obtain the most effective approach. We need to ensure the controller instances are efficiently organised to allow scalability of the overall cloud installation, while at the same time ensuring maximum security and privacy can be achieved. At this time, the Cloud Forensic Problem means that conventional cloud systems cannot guarantee GDPR compliance for cloud users. Container based solutions are likely to be subject to the same issues as conventional cloud instances. While they may very well offer some improvement, it is likely that improvement will come at an overhead cost.

Using the unikernel approach, it is likely that it will certainly be possible to be compliant with the GDPR, that the overhead of running the unikernel instances will be minimal, and that the system can be highly responsive to the need for scalability. Not only that, but the ability to provide a means for compliance for cloud systems has to be big improvement on the status quo.

While we have carried out a number of minor tests on various aspects of our proposal, we have yet to carry out any serious testing, which forms the main thrust of our next stage of the work. We have built a 'cloud in a box' with which to carry out extensive testing of our proposed system. The hardware comprises a Xeon server, running a fast Xeon processor, 16GB of fast RAM and a 525GB SSD drive, together with a 4TB fast storage drive. On this, we have loaded an HP Eucalyptus full cloud management system, which is also Amazon Web Services (AWS) compatible.

This will initially host a conventional web based system to use as a control. We will then run a system based on the proposals contained within this article. Then, we will conduct a series of typical attacks on each of the systems, and will log and analyse the results. We believe that this testing will amply support our belief that this approach will not only prove feasible, but also highly robust against attack.

Having considered how an outline technical solution might be developed, and assessing its feasibility, we now turn to the second need, namely detection of anomalous behaviour, which we address in the next section.

VIII. Detection of Anomalous Behaviour

Following the successful implementation of the solution to retaining full and proper details of the forensic and audit trails, we can now consider how we might go about detecting anomalous behaviour. Since we will now be dealing with a complete data set, then we will have a worthwhile task that we can now set about performing. Obviously, without a full forensic and audit trail available to us, it would seem rather a pointless exercise to analyse incomplete logs to attempt to detect anomalous behaviour. However, with a complete data set, this will prove to be much more worthwhile and meaningful exercise.

The common approach on this problem is often by using technical means alone. This is frequently expressed as policies authorising some action or other. However, the business architecture of an enterprise comprises a combination of people, process and technology [64], not technology alone. Such solutions are generally doomed to failure, as suggested by Duncan and Whittington in [65]–[68], who note such approaches ignore the impact of people and process on security. Both people and process are generally considered to be the weakest link in the business architecture of any enterprise.

However, in this case, we believe that to introduce people and process to the mix at this stage would be counterproductive. First, the scale of the transactional volume can be potentially enormous. Second, the work of analysis would be exceptionally boring, leading to the possibility of mistakes. Third, the introduction of people and process at this stage could lead to both errors and potential corruption, which we must consider as a large potential weakness to the system. Thanks to the robust nature of our proposed solution, we believe in this particular case, we can leave out the intervention of people and process. Naturally, the output from the system would be passed to humans for consideration and investigation, but we are confident that the analysis work on detection of anomalous behaviour could properly be performed without human intervention at this point.

We favour a straightforward approach, such as the soft security approach proposed by Neovius and Duncan [69]. In this approach, they proposed a theoretical framework that could address the highly complex challenge of securing cloud based accounting systems, which are notoriously difficult to secure properly. This would work in conjunction with an immutable database to ensure there could be no loss of audit trail or forensic records.

There is no doubt that inspecting and analysing server logs would present a very effective way to monitor what is happening with any system. Equally, there is no doubt that many companies fail to perform this rather mundane task. Usually, this comes down to a question of huge volume of transactional data, the boredom of manually analysing this data and the opportunity for errors and possible corruption due to the human input.

We suggest that leaving humans out of the main loop here would allow the work to be performed by a suitable algorithm, without the potential corrupting influence of the human input, leading to a better quality of output, performed more accurately and far more quickly. This could potentially lead to faster identification of a breach being perpetrated, thus leading to catching the culprits far more quickly and eliminating their presence from the system. Providing that appropriate forensic and audit trail data has been properly preserved, then it may be possible to ensure sufficient data is collected to assist a possible prosecution of the culprit.

There is no doubt that these tasks could also be provisioned to run on unikernels, leading to a more efficient use of resources. There is also no doubt that this is a task that cannot be left out. Analysis of server logs is one of the key ways to determine whether a breach has occurred, hopefully accompanied by sufficient forensic records to be able to do something about it. At the very least, there will be a very early warning about the possibility of an intrusion, and also the prospect of identifying what damage has been done in respect of GDPR compliance.

Providing a means of being able to identify what data has been compromised is a vital part of the armoury in mitigating the level of fines for non compliance of the GDPR. Anything we can do to ensure this can be achieved will be a good thing. Anything that can be done efficiently will be a bonus.

IX. Conclusion and Future Work

As we have already stated, the Cloud Forensic Problem presents a very serious challenge for all cloud users, especially in light of the forthcoming GDPR. We have proposed a possible solution for this problem, which is a little different from conventional approaches. However, it offers a highly robust solution to a major challenge for all organisations who will be subject to compliance with the GDPR.

We believe this solution offers such merit that we plan to run a pilot test to establish just how well it will be able to cope with a system under serious attack. Initially, it will run on a private network, under attack from professional penetration testers. Once we are sure of how well the solution is likely to perform, we will set up a real live cloud instance to see just how well it might perform.

When the GDPR comes on stream, there will not be time for organisations to mess about. If they cannot comply with the regulation, and they are breached, resulting in a loss of PII, then they can expect huge fines, the like of which they have never seen before. It is time to wake up and smell the coffee.

References

[1] B. Duncan, A. Happe, and A. Bratterud, “Using Unikernels to Address the Cloud Forensic Problem and help Achieve EU GDPR Compliance,” in Cloud Computing 2018: The Ninth International Conference on Cloud Computing, GRIDs, and Virtualization, no. February. Barcelona, Spain: IARIA, 2018, pp. 71–76.

[2] EU, “EU General Data Protection Regulation (GDPR),” 2017. [Online]. Available: http://www.eugdpr.org/ [Last accessed: August 2018]

[3] EU, “Reform of EU data protection rules,” 2016. [Online]. Available: http://ec.europa.eu/justice/data-protection/reform/index_en.htm [Last accessed: August 2018]

[4] N. Kratzke and P.-C. Quint, “Understanding cloud-native applications after 10 years of cloud computing-a systematic mapping study,” Journal of Systems and Software, vol. 126, 2017, pp. 1–16.

[5] N. Kratzke, P.-C. Quint, D. Palme, and D. Reimers, “Project cloud transit-or to simplify cloud-native application provisioning for smes by integrating already available container technologies,” Europea n Project Space on Smart Systems, Big Data, Future Internet-Towards Serving the Grand Societal Challenges. SCITEPRESS. In print, 2017.

[6] M. Almorsy, J. Grundy, and I. Müller, “An analysis of the cloud computing security problem,” 17th Asia-Pacific Software Engineering Conference (APSEC 2010) Cloud Workshop,Sydney, Australia, no. December, 2010, pp. 7.

[7] M. Almorsy, J. Grundy, and I. Mller, “An analysis of the cloud computing security problem.” The proc. of the 2010 Asia Pacific Cloud Work-shop Colocated with APSEC2010, Australia, 2010.

[8] V. Chang, Y. H. Kuo, and M. Ramachandran, “Cloud computing adoption framework: A security framework for business clouds,” Future Generation Computer Systems, vol. 57, 2016, pp. 24–41.

[9] F. Doelitzscher, T. Ruebsamen, T. Karbe, M. Knahl, C. Reich, and N. Clarke, “Sun Behind Clouds - On Automatic Cloud Security Audits and a Cloud Audit Policy Language,” International Journal on Advances in Networks and Services, vol. 6, no. 1, 2013, pp. 1–16.

[10] R. K. L. Ko, P. Jagadpramana, M. Mowbray, S. Pearson, M. Kirchberg, B. S. Lee, and Q. Liang, “TrustCloud: A Framework for Accountability and Trust in Cloud Computing,” Perspective, 2011, pp. 1–9.

[11] R. K. L. Ko, B. S. Lee, and S. Pearson, "Towards achieving accountability, auditability and trust in cloud computing," Communications in Computer and Information Science, vol. 193 CCIS, 2011, pp. 432–444.

[12] K. Lee, "Security Threats in Cloud Computing Environments," International Journal of Security and its Applications, vol. 6, no. 4, 2012, pp. 25–32.

[13] N. Papanikolaou, S. Pearson, and M. C. Mont, "Towards Natural-Language Understanding and Automated Enforcement of Privacy Rules and Regulations in the Cloud: Survey and Bibliography," Analysis, 2011, pp. 1–9.

[14] S. Pearson, "Taking Account of Privacy when Designing Cloud Computing Services," Current, 2009, pp. 44–52.

[15] S. Pearson, "Toward accountability in the cloud," IEEE Internet Computing, vol. 15, no. 4, jul 2011, pp. 64–69.

[16] S. Ramgovind, M. Eloff, and E. Smith, "The Management of Security in Cloud Computing," in Information Security for South Africa (ISSA), 2010, 2010, pp. 1–7.

[17] H. Takabi, J. B. Joshi, and G.-J. Ahn, "Security and Privacy Challenges in Cloud Computing Environments," IEEE Security & Privacy Magazine, vol. 8, no. 6, nov 2010, pp. 24–31.

[18] S. Creese, P. Hopkins, S. Pearson, and Y. Shen, "Data Protection-Aware Design for Cloud Computing," Work, no. December, 2009, pp. 1–13.

[19] K. Hon, C. Millard, and I. Walden, "The Problem of 'Personal Data' in Cloud Computing - What Information is Regulated ?" 2011.

[20] W. K. Hon, C. Millard, and I. Walden, "Who is Responsible for 'Personal Data' in Cloud Computing? The Cloud of Unknowing, Part 2," Legal Studies, no. 77, 2011, pp. 1–31.

[21] W. K. Hon, J. Hörnle, and C. Millard, "Data Protection Jurisdiction and Cloud Computing When are Cloud Users and Providers Subject to EU Data Protection Law?" Legal Studies, 2011, pp. 1–40.

[22] W. Itani, A. Kayssi, and A. Chehab, "Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing Architectures," 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, dec 2009, pp. 711–716.

[23] W. Jansen and T. Grance, "Guidelines on Security and Privacy in Public Cloud Computing," NIST, Tech. Rep., 2011. [Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf [Last accessed: August 2018]

[24] H. Katzan Jr, "On The Privacy Of Cloud Computing," International Journal of Management and Information Systems, vol. 14, no. 2, 2011, pp. 1–12.

[25] W. K. Hon, C. Millard, J. Singh, I. Walden, and J. Crowcroft, "Policy, legal and regulatory implications of a Europe-only cloud," International Journal of Law and Information Technology, vol. 24, no. 3, 2016, pp. 251–278.

[26] NIST, "Security and Privacy Controls for Federal Information Systems and Organizations Security and Privacy Controls for Federal Information Systems and Organizations," Natioinal Institute of Standards and Technology, Gaithersburg, MD, Tech. Rep. February, 2014. [Online]. Available: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf [Last accessed: August 2018]

[27] S. Pearson and A. Charlesworth, "Accountability as a Way Forward for Privacy Protection in the Cloud," Computing, no. December, 2009, pp. 1–15.

[28] S. Pearson and A. Benameur, "Privacy, Security and Trust Issues Arising from Cloud Computing," 2010 IEEE Second International Conference on Cloud Computing Technology and Science, Nov 2010, pp. 693–702.

[29] S. Pearson, "Privacy, Security and Trust in Cloud Computing," in Privacy and Security for Cloud Computing. e: Springer, 2013, pp. 3–42.

[30] J. Prüfer, "How to govern the cloud? Characterizing the optimal enforcement institution that supports accountability in cloud computing," Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom, vol. 2, 2013, pp. 33–38.

[31] S. S. Shapiro, "Privacy by Design," Communications of the ACM, vol. 53, no. 6, jun 2010, p. 27.

[32] J. Singh, T. F. J. M. Pasquier, and J. Bacon, "Securing tags to control information flows within the Internet of Things," 2015 International Conference on Recent Advances in Internet of Things, RIoT 2015, 2015.

[33] EU, "Accountability for Cloud (A4Cloud)," 2018. [Online]. Available: http://a4cloud.eu/ [Last accessed: August 2018]

[34] C. A. Adams and R. Evans, "Accountability, Completeness, Credibility and the Audit Expectations Gap," JCC 14 Summer 2014, vol. 14, no. Summer, 2014, pp. 97–115.

[35] W. Benghabrit, H. Grall, J.-C. Royer, M. Sellami, M. Azraoui, K. Elkhiyaoui, M. Onen, A. S. D. Olivera, and K. Bernsmed, "A Cloud Accountability Policy Representation Framework," in CLOSER-4th International Conference on Cloud Computing and Services Science, 2014, pp. 489–498.

[36] K. Bernsmed, W. K. Hon, and C. Millard, "Deploying Medical Sensor Networks in the Cloud Accountability Obligations from a European Perspective," in Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on. IEEE Comput. Soc, 2014, pp. 898–905.

[37] D. Butin, M. Chicote, and D. Le Métayer, "Log design for accountability," in Proceedings - IEEE CS Security and Privacy Workshops, SPW 2013, 2013.

[38] D. Catteddu, M. Felici, G. Hogben, A. Holcroft, E. Kosta, R. Leened, C. Millard, M. Niezen, D. Nunez, N. Papanikolaou, S. Pearson, D. Pradelles, C. Reed, C. Rong, J.-C. Royer, D. Stefanatou, and T. W. Wlodarczyk, "Towards a Model of Accountability for Cloud Computing Services," in International Workshop on Trustworthiness, Accountability and Forensics in the Cloud (TAFC), 2013, pp. 21–30.

[39] A. Haeberlen, "A Case for the Accountable Cloud," ACM SIGOPS Operating Systems Review, vol. 44, no. 2, 2010, pp. 52–57.

[40] W. Hon, E. Kosta, C. Millard, and D. Stefanatou, "Cloud Accountability: The Likely Impact of the Proposed EU Data Protection Regulation," Queen Mary School of Law Legal Studies Research Paper, no. 172, 2014, pp. 1–54.

[41] K. L. R. Ko, P. Jagadpramana, and B. S. Lee, "Flogger: A File-Centric Logger for Monitoring File Access and Transfers within Cloud Computing Environments," Computing, 2011, pp. 1–8.

[42] T. Lynn, P. Healy, R. McClatchey, J. Morrison, C. Pahl, and B. Lee, "The Case for Cloud Service Trustmarks and Assurance-as-a-Service," in CLOSER 2013 - Proceedings of the 3rd International Conference on Cloud Computing and Services Science, 2013, pp. 110–115.

[43] N. Papanikolaou, S. Pearson, M. C. Mont, and R. K. L. Ko, "Towards Greater Accountability in Cloud Computing through Natural-Language Analysis and Automated Policy Enforcement," Engineering, 2011, pp. 1–4.

[44] N. Papanikolaou, T. Rübsamen, and C. Reich, "A Simulation Framework to Model Accountability Controls for Cloud Computing," CLOUD COMPUTING 2014, The Fifth International Conference on Cloud Computing, GRIDs, and Virtualization, no. c, 2014, pp. 12–19.

[45] S. Pearson, M. C. Mont, and G. Kounga, "Enhancing Accountability in the Cloud via Sticky Policies," in Secure and Trust Computing, Data Management, and Applications, 2011, pp. 146–155.

[46] S. Pearson, V. Tountopoulos, D. Catteddu, S. Mario, R. Molva, C. Reich, S. Fischer-Hubner, C. Millard, V. Lotz, M. G. Jaatun, R. Leenes, C. Rong, and J. Lopez, "Accountability for Cloud and Other Future Internet Services," in CloudCom, 2012, pp. 629—-632.

[47] K. Bernsmed and S. Fischer-Hübner, "Secure IT Systems: 19th Nordic Conference, NordSec 2014 Tromsø, Norway, October 15-17, 2014 Proceedings," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8788, 2014, pp. 3–24.

[48] T. Ruebsamen and C. Reich, "Supporting Cloud Accountability by Collecting Evidence Using Audit Agents," in CloudCom 2013, 2013, pp. 185–190.

[49] J. Singh, J. Powles, T. Pasquier, and J. Bacon, "Data Flow Management and Compliance in Cloud Computing," Cloud Computing, no. Special Issue on Legal Clouds., 2015, pp. 1–12.

[50] A. Squicciarini, S. Sundareswaran, D. Lin, and S. Huang, "Promoting Distributed Accountability in the Cloud," in IEEE 4th International Conference on Cloud Computing Promoting, 2011, pp. 113–120.

[51] S. Sundareswaran, A. C. Squicciarini, and D. Lin, "Ensuring Distributed Accountability for Data Sharing in the Cloud," in IEEE Transactions on Dependable and Secure Computing, vol. 9, no. 4, 2012, pp. 556–568.

[52] M. Theoharidou, N. Papanikolaou, S. Pearson, and D. Gritzalis, "Privacy risk, security, accountability in the cloud," in Proceedings of the

International Conference on Cloud Computing Technology and Science, CloudCom, vol. 1, 2013, pp. 177–184.

[53] Verizon, "2012 Data Breach Investigation Report: A study conducted by the Verizon RISK Team in cooperation with the United States Secret Service and Others," Tech. Rep., 2012. [Online]. Available: http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2012_en_xg.pdf [Last accessed: August 2018]

[54] Verizon, "2016 Verizon Data Breach Report," Tech. Rep., 2016.

[55] A. Bratterud, A.-A. Walla, H. Haugerud, P. E. Engelstad, and K. Begnum, "IncludeOS: A Minimal, Resource Efficient Unikernel for Cloud Services," 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom), 2015, pp. 250–257.

[56] N. Kratzke, "About microservices, containers and their underestimated impact on network performance," arXiv preprint arXiv:1710.04049, 2017.

[57] B. Duncan, A. Bratterud, and A. Happe, "Enhancing Cloud Security and Privacy: Time for a New Approach?" in Intech 2016, Dublin, 2016, pp. 1–6.

[58] A. Bratterud, A. Happe, and B. Duncan, "Enhancing Cloud Security and Privacy: The Unikernel Solution," in Cloud Computing 2017: The Eighth International Conference on Cloud Computing, GRIDs, and Virtualization, 2017, pp. 1–8.

[59] A. Happe, B. Duncan, and Alfred Sewitsky Bratterud, "Unikernels for Cloud Architectures: How Single Responsibility can Reduce Complexity, Thus Improving Enterprise Cloud Security," in COMPLEXIS 2017 - Proceedings of the 2nd International Conference on Complexity, Future Information Systems and Risk, Porto, Portugal, 2017, pp. 1–12.

[60] B. Duncan, A. Happe, and A. Bratterud, "Enterprise IoT Security and Scalability: How Unikernels can Improve the Status Quo," in 9th IEEE/ACM International Conference on Utility and Cloud Computing (UCC 2016), Shanghai, China, 2016, pp. 1–6.

[61] B. Duncan and M. Whittington, "Creating an Immutable Database for Secure Cloud Audit Trail and System Logging," in Cloud Computing 2017: The Eighth International Conference on Cloud Computing, GRIDs, and Virtualization. Athens, Greece: IARIA, ISBN: 978-1-61208-529-6, 2017, pp. 54–59.

[62] B. Duncan and M. Whittington, "Creating and Configuring an Immutable Database for Secure Cloud Audit Trail and System Logging," International Journal On Advances in Security, vol. 10, no. 3&4, 2017, pp. 155–166.

[63] OWASP, "OWASP home page," 2017. [Online]. Available: https://www.owasp.org/index.php/Main_Page [Last accessed: August 2018]

[64] PWC, "UK Information Security Breaches Survey - Technical Report 2012," PWC2012, Tech. Rep. April, 2012.

[65] B. Duncan, D. J. Pym, and M. Whittington, "Developing a Conceptual Framework for Cloud Security Assurance," in 2013 IEEE 5th International Conference on Cloud Computing Technology and Science. Bristol: IEEE, 2013, pp. 120–125.

[66] B. Duncan and M. Whittington, "Compliance with Standards, Assurance and Audit: Does this Equal Security?" in Proceedings of the 7th International Conference on Security of Information and Networks. Glasgow: ACM, 2014, pp. 77–84.

[67] B. Duncan and M. Whittington, "The Importance of Proper Measurement for a Cloud Security Assurance Model," in 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom), Vancouver, 2015, pp. 1–6.

[68] B. Duncan and M. Whittington, "Information Security in the Cloud: Should We be Using a Different Approach?" in 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom), Vancouver, 2015, pp. 1–6.

[69] M. Neovius and B. Duncan, "Anomaly Detection for Soft Security in Cloud based Auditing of Accounting Systems," in Closer 2017 - 7th International Conference on Cloud Computing and Services Science, Porto, Portugal, 2017, pp. 1–8.
[Last accessed: August 2018]