Survey of Cloud Server Geolocating Techniques

Leo Hippelainen[1, 2], Ian Oliver[1], Shankar Lal[1, 2]
[1]Nokia Bell Labs, Espoo, Finland
[2]Aalto University, Espoo, Finland
[1]first.last@nokia-bell-labs.com, [2]first.last@aalto.fi

***Abstract*—Every cloud data center has a geographical location, which is known to, at least, its owner. Still, it is a challenge for a cloud customer to discover indisputably the location of the data center, which hosts an active cloud based service and its data. Knowing the location is necessary to assess that pertinent geographical obligations of service level agreements and legal requirements are fulfilled. For example, several countries have legislation that denies processing private data of its citizens in countries that do not have compliant legislation. This paper presents existing technologies that can be used to resolve and assess the geographical location of a cloud servers.**

I. Introduction

Knowing geographical location of a cloud data center has become a significant question to the cloud customers [1] and privacy enforcement authorities [2]. Some use cases, like processing governmental information or storing private information of citizens, are impacted by national laws that may require processing and saving data to happen only within domestic territory. Violations may become expensive to enterprises: for example, General Data Protection Regulation (GDPR) [3] may yield a fine up to 20 million EUR or 4% of the total worldwide annual turnover, whichever is higher.

In contrast, one of the characteristics of cloud computing is smooth migration of computing and storage workload within and among the data centers. With ordinary cloud based services this is not an issue: service level agreement (SLA) terms become fulfilled and customer gets what she/he has subscribed for.

When cloud based computing is becoming more widely applied and worries about privacy and confidentiality concerns increase, a contradiction emerges between cloud provider's interest to relocate computing resources according to his/her interests and some cloud customer's interest to comply with legal restrictions, which impose limits to geographical location of data processing and storage [4].

A classic example of the requirements of geolocation related to service provision and data storage can be found in the case of Lawful Intercept (LI) where access to data is granted based on a Judge's (or similar authority) orders. In this case access and subsequent storage and processing of data is wholly within a nation's physical borders. For telecommunications operators of all kinds this places strict requirements on the physical placement of computing, including virtual, resources.

In ideal case a cloud customer can fully trust that their cloud provider is honest and open about where their servers are located and adhere strictly to the terms of the SLAs. However, in case of any doubt, trustful, real-time verification of the claimed geographical locations by the customer becomes challenging. It can be next to impossible if the provider is not willing to co-operate or cannot be trusted.

Awareness about location may become one of the essential requirements that drives selection of cloud service provider. Concerned customers will create market pressure to cloud providers for supporting better transparency of geographical location of the offered resources.

Geographical location of data is important because it determines which legislation is applied. Landing data on foreign territory may open route to curious foreign actors or other possibly malicious parties to access sensitive data.

On the other hand, possibility to data leaks may also be used as an excuse to justify protective policies and laws in favor of national data center operators. In any case, the location of data must be undeniably attested.

Data records are often stored to several locations during their lifetime. They have not only the primary storage location but also data replica locations, and at processing time there are copies in run-time memory and CPU registers. Moreover, data gets transmitted between computer system devices and between data centers. If all these data appearances occur within a geographical area, which has sufficient legislation, and if data still gets stolen, we at least should be able to prosecute the wrongdoer. Obviously we must also be able to detect and proof the breach and find out the suspect.

Data breaches go often unnoticed for a long time. Those that are detected may give clues that the attackers are supported by foreign governmental actors. This challenges common assumption in the literature about economic rationality of the attackers. When the attackers is not driven by business rationales, they can invent attack patterns which otherwise could be ignored.

Indoor locating techniques have been widely reported in the state-of-the-art. Majority of the research aims to finding out the location coordinates of mobile devices. When locating stationary objects, like data center server hardware, same technologies can be used but the offered products are typically too feature rich and unnecessary expensive for our purpose. We are not interested to locate a certain server within the site, we just want to know if it is there or not.

The rest of this paper is organized as follows: Section II introduces terms and concepts that are necessary to understand the domain. Section III describes constraints and requirements to limit the problem domain. Section IV presents several technical algorithms and approached that could be useful in providing solutions. Section V scans most important legislation worldwide that cover transborder data flows. Section VI evaluates applicability of potential technologies to the problem domain. Finally, Section VII summarizes the work and outlines future efforts.

II. Terms and Technologies

We will introduce essential concepts to deal with the problem area. They include basic entities, stakeholders, cloud service and deployment models and cloud server roles.

A. *Basic entities*

In this section we list the very basic definitions of a data center.

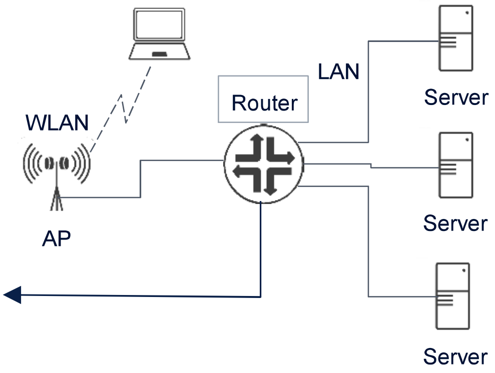

Fig. 1 Simplified Data Center Elements

1) Data Center: A data center is a facility used to house computer systems and associated components, such as telecommunications and storage systems, backup power supplies, redundant data communications connections, environmental controls (e.g., air conditioning, fire suppression) and various security devices [5]. For the purpose of this paper we can simplify a data center to denote a set of interconnected servers and storage systems with communication connections to the outside world.

At any given time a data center has a geographical location. The location implies the jurisdiction of the data center.

2) Cloud Computing: Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models [6].

3) Cloud Data Center (CDC): The kind of a data center that hosts computer systems suitable for use in cloud computing.

4) Local Area Network (LAN): A fast data transmission network to interconnect computers in a small local area, like a data center. Each computer has a physical adapter, a network interface card (NIC), which connects the computer to a LAN router or switch. A computer can have one or several NICs that are connected to the same or different LANs.

5) Router:

Cross-connection of LAN or WLAN lines. A router forwards incoming data packets to outgoing ports based on the destination address in the data packet header. Usually a LAN has several interconnected routers and other network equipment.

6) Server: A computer with processor cores, random access memory, LAN connectivity and control circuitry. Often a server also has mass memory (e.g., disk drive), a display port and sockets for options, like Trusted Platform Module (TPM) chip. More about server categories in the next subsection.

7) Wireless Local Area Network (WLAN): Similar to LAN but uses wireless data links provided by Access Points (AP) and wireless network transceivers in the connected devices.

B. *Cloud server categories*

Cloud data center servers can be sorted to three groups: compute, storage, and control servers. Cloud Customer's applications are running in compute servers and they access Cloud Customer's data via storage servers. Control servers are used for managing the cloud system. Typically there are special server hardware variants optimized for each role.

1) Compute Server: Compute servers execute platform software, like, middleware, and also Cloud Customer's applications. They contain temporary copies of data when processing it as per need by a cloud service applications. Therefore the geographical site of the compute servers needs to be managed and known.

2) Storage Server: The purpose of storage servers is to host data reliably on its mass memory devices. Storage servers do not execute Cloud Customer's applications, but instead platform software, like database engines. The applications invoke functionality published by the platform. Location of the storage servers must be known and shared with an interested cloud customer.

3) Control Server: Control servers do not process Cloud Customer's data, but they schedule compute and storage servers for use by Cloud Customer's application and platform software. Hence control servers must be aware of location to the extent it affects scheduling. Locations of the control servers themselves are not interesting from our perspective.

C. *Stakeholders*

There are various parties who are interested in knowing the geographical location of cloud servers.

1) Cloud Customer: Cloud Customer employs compute and storage resources made available by one or several Cloud

Providers for executing cloud based services implemented by the customer's application software. Cloud Customer may have legal obligations to provide the service using servers within a certain legislation.

2) Cloud Provider: Cloud Provider makes available a set of compute and data storage servers to host application software. The hardware can be in data centers at different geographical locations on the globe. Cloud Provider wants to maximize profit and sales of his/her cloud resources by offering competitive value for the money to Cloud Customers. Cloud Provider also wants to share computing workload optimally to the servers across his/her data center network.

3) Cloud Service End User: A human or an automated actor who consumes functions of a service provided by Cloud Customer. This role is not important from the viewpoint of this paper but it is mentioned here to emphasize that there are, at least, two levels of customers to cloud services.

4) Data Protection Officer (DPO): DPO is a person with expert knowledge of data protection laws and practices. He/she should assist the data controller (Cloud Customer) or processor (Cloud Provider) to monitor internal compliance with GDPR [3] or equivalent law. DPO will be under a legal obligation to notify the Supervisory Authority without undue delay about data breaches.

5) External auditor: External auditor is an independent auditor trusted by both Cloud Provider and Cloud Customer whom are in a business relationship. External Auditor can perform formal audits to a data center and produce audit reports about conformance of Cloud Provider's services against applicable laws, regulations and SLAs. External Auditor appreciates easiness of auditing geographical locations of physical servers of Cloud Provider.

D. Cloud service models

Commonly used cloud service models are defined by NIST [6]. Fig. 2 depicts responsibilities of Cloud Provider and Cloud Customer with the service models.

1) Infrastructure as a Service (IaaS): Cloud Provider is responsible for the data center infrastructure and software stack up to and including the hypervisor, which maintains virtualization. The customer has control over operating systems, programming frameworks, middleware, and deployed applications including their data; and possibly limited control of select networking components (e.g., host firewalls).

2) Platform as a Service (PaaS): Cloud Provider offers also programming frameworks, libraries, services, and tools. The customer has control over the deployed applications, their data and possibly configuration settings for the application-hosting environment.

3) Software as a Service (SaaS) Cloud Provider takes care of the whole software stack. The customer may have control over user-specific application configuration settings.

E. Cloud deployment models

Cloud computing can be arranged in various combinations.

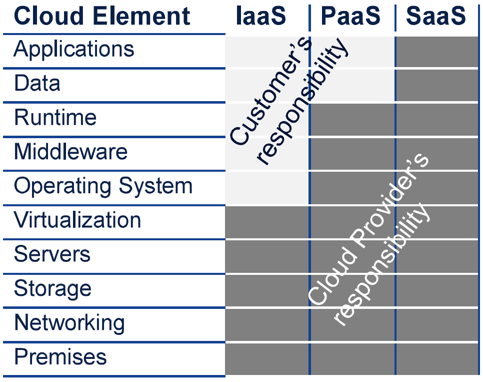

Fig. 2 Responsibilities in Cloud Service Models

Common deployment models are as explained next [6], [7].

1) Private Cloud: The cloud infrastructure is provisioned for exclusive use by a single organization. It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

*2) Public Cloud.*The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider

3) Community Cloud: The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns. It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

4) Hybrid Cloud: The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns. It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

III. Problem Description

In this section, we define our research domain and challenges related to geographical location of cloud servers.

A. Assumptions and requirements

Our primary concern is to validate in an undisputable manner that a physical server exist in a certain data center site. We are not interested in knowing the server's exact location within the data center.

1) Cost Awareness: Price of a server is assumed to be in range of 1,000 to 10,000 euros. We assume that it should be acceptable to allow 10 euros or so increase to the manufacturing costs of a server to add necessary hardware to the motherboard for achieving undeniable locating functionality.

2) Reliability: False positives, i.e. detecting a server to exist at a site when it actually does not, shall not happen, because our main objective is to detect servers which are not at the site they claim to be. False negatives should not happen, but if there are such incidents, the consequence is that the geolocation information of the server in question is not regarded as trusted. This mistake is not severe as it does not cause location critical workload to be scheduled to an untrusted server. It can initiate a process to clarify the actual location in case the server was earlier considered as having geological trust. An acceptable rate of false negatives is obviously use case specific. For example, it could be 0.1%. But of course, the smaller the better.

3) Volume: A data center site can have thousands or even millions physical servers and a cloud system may extend to several data center sites. Geographical location should be one of the properties based on which the cloud computing resources are allocated for running the services of the customers.

4) Radio Signal Propagation: A data center site is expected to be protected against electromagnetic disturbances by its walls acting as a Faraday cage or it can be underground deep in a cave. Consequently, we cannot assume that radio signals from outside the data center will reach antennas inside. On the other hand, we cannot assume that strong enough signals originating from inside the data center cannot reach antennas outside.

5) Geographical Location: A data center site is expected to reside at least few kilometers from the border of a jurisdiction.

6) Auditing: External Auditor can visit a data center site and personally verify, for example, that a certain server hardware exists at that site.

7) Server Identifiers: Each server has a unique identifier, like serial number, which remains the same during lifetime of the server. The identifier must be available at run-time to software. If the identifier is changed, the server is considered to be a different one.

8) Mobility: Physical servers are not expected to change data center site very often after being commissioned to a site. Nevertheless, it can still happen a couple of times during the lifetime of a server.

It is possible that a data center is assembled in a sea container [8] or several of them. Technically it is possible that the whole data center is moved to another geographical area, also to another jurisdiction without being reconfigured.

B. Research problems

In this section we outline viewpoint in the domain and derive specific problems that need resolved. These problems are used as guidelines for introducing technologies that can be helpful while finding the solutions.

1) Cloud Provider optimizes utilization of server resources: An honest Cloud Provider is willing to comply with constraints agreed with Cloud Customer, which may include limitations to the location of data centers that can host a certain service. How the cloud resource scheduler should respect location constraints when allocating resources to hosted applications?

2) Cloud Customer verifies that Cloud Provider respects geographical constraints: Cloud Customer has promised to its End User that certain geographical limitations are respected when assigning cloud resources to his/her application. Cloud Customer who can proactively ascertain that geographical limitations are indeed respected, gains better reputation among its potential customers and thus can win more business. Consequently, Cloud Providers that support assessing geographical location dependably at run-time improve their position among Cloud Customers.

How to facilitate trustworthy location information concerning the allocated physical servers when Cloud Customer's services are up and running?

How to find out where the data records are stored?

3) End user wants to verify that his/her data is not ported to uncompliant jurisdiction. Privacy of End User's data records (e.g., sensitive personal information, lawful interception call records, national security plans) and data sovereignty (explained in section V) combined are the driving force behind concerns about geographical location of employed cloud resources. It can be that Cloud Customer, while executing service, accumulates metadata (e.g., directory of call recordings) which becomes also as location discrete as the actual data records.

How can an end user dependably verify the geographical locations of physical resources being used for handling his/her data?

4) External Auditor wants to check locations of cloud servers: If there is no trust between Cloud Provider and Cloud Customer or there is not enough legally valid evidence of it, an External Auditor can be contracted to produce necessary testimonial about geographical locations of servers. This delegates the challenge of acquiring dependable location information to External Auditor.

How to audit, with reasonable effort, a data center containing thousands of servers and assure location of each physical server?

External Auditor can testify situation at the moment of the audit. How to make Cloud Customers notified if the situation changes after an audit?

After the locations of the cloud resources are verified, how to assess that the location information is indeed respected while allocating resources even during busy moments when there might be shortage of available resources?

5) Cheating patterns of a dishonest Cloud Provider: To deserve trust in the provided solutions we must be prepared to tackle by-passes that a dishonest Cloud Provider can come up with to cheat Cloud Customers and auditors concerning geographical locations of physical servers.

Cloud Customer employers typically never visit the data center site an even if they could, they have no expertize to make judgement if compute and storage resources they sell to their End Users are actually allocated from the allowed site group. Thus they must trust to earlier performed audits and possible technical means to attest the location.

How to validate evidence provided by Cloud Provider concerning geographical location of a virtual server?

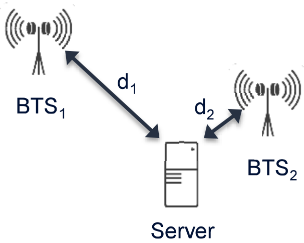

Fig. 3 Simple locating setup

How to get alarm or at least find out afterwards that geographical limitations are or were respected?

The above questions can be applied also to honest Cloud Providers who have vulnerabilities in their system and there are malicious third party attacks which utilize those vulnerabilities.

6) Cheating patterns of a dishonest Cloud Customer: Dishonest Cloud Customer can promise location awareness to its End Users but instead decides to increase its profit by subscribing resources from a bulk cloud provider that does not promise location based allocation.

Similarly to Cloud Customer employees also End User employees must trust on the evidence the Cloud Customer can offer, perhaps with help of the Cloud Provider. For example, the evidence could be collected using software running in the same cloud as the application. There should exist means to validate the evidence from independent sources.

IV. LOCATION ALGORITHMS

Location detection algorithms relevant to the research question here can be categorized based on their algorithms, such as:

- Proximity to a transceiver
- Signal strength
- Signal delay
- Signal direction
- Distance-bounding protocols
- IP address based mapping
- Server naming
- Provisioned location code
- Visual image
- Network Topology
- Planetary constants
- Satellite based positioning systems
- Combination Techniques
- Attestation service

Each of these will be covered by following subsections. Each alternative is evaluated from the viewpoint of using it for locating physical servers in a data center. Literature [9], [10] mentions also some other methods, but from our viewpoint they do not offer significant advantages compared to the selected ones.

A. *Proximity to a transceiver*

The simplest method of locating a target device using a transmitted signal is to define its location to be the same as that of the nearest base transceiver station (BTS) of the communication system [9]. Various standards call BTS with different names: Bluetooth as beacon, ZigBee as static node (SN), Wi-Fi as access point (AP), 2nd generation (2G) cellular networks as base transceiver station (BTS) and 3G networks as Node B (NB) or Evolved Node B (eNB).By "nearest" we mean here the one with the best signal reception. Location is unknown if there is no signal from any BTS.

In Fig. 3 the signal from Server to BTS2 travels shorter distance than to BTS1. Consequently, BTS2 should get a stronger signal and thus the server should be considered to exist at the coordinates of BTS2. In our case a more typical use for proximity is when there is only one BTS within reach of a server. This setup is enough to detect if a server is in that data center site or not, which is the information we actually want to know.

Proximity based locating is not ideal for pinpointing location accurately. Even in case of several BTSs we just want to know if any of the BTSs can receive signal from the target device.

B. *Radio Signal Strength*

More accurate information about the location of a device that what proximity can provide, we can measure signal strength received at several nearby BTSs. We can assume that received signal strength is a function of distance. If we can get Radio Signal Strength Indicator (RSSI) reading from at least three BTSs, we can compute using trilateration, how far the target device resides from to the BTSs [11]. In ideal case there can be only one spot where the distances match observed signal strengths.

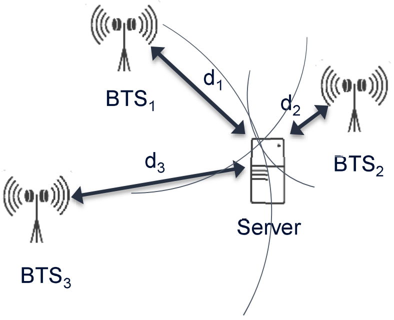

Fig. 4 Trilateration Measurements

To facilitate locating the target device, we must know location coordinates of every BTS. A typical implementation can be a database, which relates BTS identifier with certain geographical location. Assuming that signal strength is a function of distance, we can use trilateration algorithm to compute the location.

Locating accuracy of this approach depends on the geographical distribution of BTSs in the area, used radio spectrum and transmission power, and distribution of radio wave obstacles and reflectors. In some frequency ranges the signal strength may fluctuate strongly even with small movement distances because of interference. This causes error to the computed location information. From the server locating point of view roaming in not an issues similarly than with mobile target devices.

C. *Signal Direction*

Another method for locating is to measure the angle of direction from which a signal arrives. If the angle is measurable from two BTSs, then the location can be computed using triangulation [9], [11].

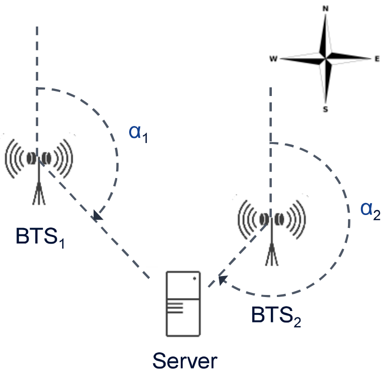

Fig. 5 Triangulation Measurements

Fig. 5 shows symbolically two base stations. The angles α_1 and α_2 are measured as compass directions on horizontal plane.

Typical method for measuring angle uses array of antennas, in which each antenna covers a narrow sector. Absolute inaccuracy increases with the distance between the BTS and the target device. Also if the strongest from the device is a reflection, locating algorithm will give wrong result.

D. *Signal Delay*

Distance measurement can also be based on the time the signal travels between a target device and BTSs [9] because the travel time is a linear function of the distance (Fig. 6). If the device and the BTSs have common time base, we can measure Time Of Arrival (TOA), if not, Time Difference Of Arrival (TDOA). The time difference refers to arrival times measured at BTSs, which must be time synchronized.

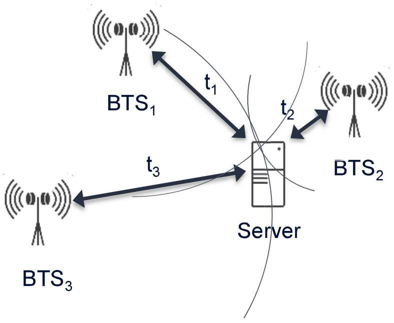

Fig. 6 Time of Arrival Measurements

Time Division Multiple Access (TDMA) based cellular systems, like GSM, use Timing Advance parameter to compensate signal delay from distant mobile so that signal burst sent by it arrives within the reserved time slot [12]. Consequently, every GSM BTS knows the distance to every mobile phone registered to it with accuracy of about 0.5km.

Locating accuracy of this method depends on the accuracy of time measurements. Reflections can cause inaccuracy if the received signal is not arriving through the shortest path.

E. *Distance-Bounding Protocols*

A distance-bounding protocol is an authentication protocol between a verifier V and a prover P, in which V can verify the claimed identity and physical location of P. This protocol may be used to monitor the geographic location of the remote server by measuring the physical distance by timing a round trip time (RTT) between sending out challenge bits and receiving back the corresponding response bits [13].

Transmission latency measurement has the same basic idea as signal delay measurement, but instead of measuring propagation time of the signal, we measure transmission delay in the computer network by sending a polling request and waiting for response to it. RTT algorithm can also be applied to local positioning using Wi-Fi network [14].

In global network the method requires a set of landmark computers at known locations, which are used for measuring reference latencies. Assuming the computer network is homogenous with similarly behaving routers and transmission links, transmission latency between two computers correlates with their geographical distance. If we have three polling points with known distances to landmarks, we can compute first probable distances from the measurement points to a server under investigation and then approximate global position of the server using trilateration method.

F. *IP address based mapping*

Domain registration records, accessible through whois databases, provide the physical address of the registrant for an IP address. However, there are no guarantees that these records have accurate information due to lazy updates and what the registrant has announced. An adversary can also dynamically reassign an IP address [15].

Traceroute discovers the path packets traverse using ICMP or TCP. Unfortunately, an intermediate router may drop these packets leaving only a partial route. The real problem with traceroute and 'whois' services is that they use an IP address as input. Colluding storage nodes have the ability to send packets with the same IP address. As a result, a distant machine could respond to storage challenges even though the IP address appears to be close.

G. Provisioned location code

The geographical location can be provisioned to the server when it is commissioned for use [4], [16]. This is an accurate method as long as the location information is updated whenever the server is relocated. We must also trust to whoever is responsible for configuring correct location information.

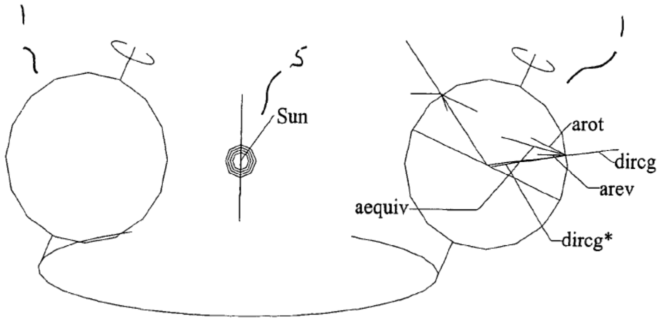

Fig. 7 Planetary acceleration vectors

One of Trusted Platform Module (TPM) registers can be reserved (often PCR 22) to store geographical location information.

H. Visual image

Visual images can be used for locating objects in two different manners: the object is pinpointed from a surveillance camera picture or the object sends pictures of its surroundings, which are then compared to pictures of known locations.

Server rooms of a data center may have surveillance cameras. If each server is visible in some surveillance camera, basically we can use this visual evidence to attest existence of a server in a certain data center. With help of software, which can blink a front panel light upon command, the auditor can verify from a live video stream if the physical server is in the picture. However, a dishonest cloud service provider can easily hack the blinking program to mislead the auditor.

Concerning the second approach, having video camera installed to every server seems like an expensive solution. Even if economically feasible, surroundings of servers and, consequently, view seen by cameras, may not differ enough to facilitate reliable evidence that a server is in a data center. However, we can have an independent mechanism, like turning lights on in the server room, which can be triggered by the auditor, to gather evidence that a server is in a certain data center. Also this scenario offers dodges for a dishonest provider.

I. Server naming

A recent study [17] researchers present geographical locations of a worldwide video-streaming service provider (Netflix). Their method was to watch films using client computer in different parts of the world and observing the server's IP address and name. It turned out that the server names followed a certain naming pattern, in which the geographic location was hinted as the nearest airport code.

Obviously locating method based on server names is not dependable because there is no guarantee that the names are set honestly.

J. Network Topology

As already mentioned above, geographical location does not necessarily correlate with computer logical network topology. Nevertheless, data centers typically used wired LAN to connect the servers. By following systematic practices, location within data center can be derived from the data switch and its port, into which a NIC of a server is wired. However, this is not an interesting locating problem for us, who are not looking for ways to know the data center site of a physical server.

K. Planetary Constants

US patent 7,822,549 [18] describes a global locating method based on measuring the vector sum of the centripetal acceleration of the rotation of the Earth around its axis and the in its orbit around the Sun. The measurements require sensitive instruments and some time. Anyway, this method would produce an independent estimate of the geographical location.

This method apparently has not been a commercial success in the data center domain probably due to expensive price of adding sensitive enough accelerometers to every physical server.

L. Satellite Based Positioning Systems

Global Positioning System (GPS) and other similar global navigation satellite systems are commonly used for outdoor location detection. However, our assumption that radio signals are not guaranteed to enter data center premises, exclude possibility of using satellite based systems. There are also spoofing devices which can feed misleading locating signal to GPS (and alike) receivers [19].

M. Combination Techniques

Several independent locating techniques can be used together to provide more accurate results. For example, Google, Microsoft, and other companies maintain database containing geographical coordinates of Wi-Fi APs [20]. They collect the location data by using cell phones, which have GPS capability. The phone is made periodically to check, which BTSs and APs are in the neighborhood and what is their signal strength. This data can be used to make the location database more accurate.

With help of the collected database it is possible to locate a mobile terminal much faster than relying on satellite based positioning system alone. It also supports locating indoors, where GPS (or equivalent) signal is not usually available.

N. Attestation Service

We could have an attestation server that has a database of all physical cloud servers and which has certified geolocation data. The server must be outside direct control of Cloud Providers. It uses universally unique identifier of each server to index the corresponding geographical site and location. In addition there must be a commonly agreed and trusted application programming interface (API) to retrieve the unique identifier at run-time from an application in a virtual machine.

This approach seems to resolve the location detecting problem. Nevertheless, there are some issues: the server location database becomes a critical resource and bottleneck. The real problem is to maintain server location database in real-time and automatically update said data this the reader will note is the research problem being addressed.

Location data of every server is private data of Cloud Providers, which implies that it should be stored to data centers in trusted jurisdictions. Also maintenance of the database may become an issue, especially if the Cloud Provider wants to cheat. We loop back to the root problem of this paper: How to assure that server's geolocation data is dependable?

V. Data Residency and Sovereignty Legislation

Data sovereignty is the concept that information, which has been converted and stored in binary digital form, is subject to the laws of the country in which it is located. Many of the current concerns that surround data sovereignty relate to data that is stored in a foreign country from being subpoenaed by the host country's authorities or some malicious actors, because prevalent laws do not set prohibiting enough punishments or because monitoring is not strong enough [21].

Data residency refers to the physical or geographic location of an organization's data or information. Similar to data sovereignty, data residency also refers to the legal or regulatory requirements imposed on data based on the country or region in which it resides [22].

The wide-spread adoption of cloud computing services, as well as new approaches to data storage including object storage, have broken down traditional geopolitical barriers more than ever before. In response, many countries have regulated new compliance requirements by amending their current laws or enacting new legislation that requires customer data to be kept within the country the customer resides [21], [23].

Verifying that data exists only at allowed locations can be difficult. It requires the cloud customer to trust that their cloud provider is completely honest and open about where their servers are hosted and adhere strictly to service level agreements (SLAs) [21].

According to OECD [24] over sixty countries had adopted by year 2011 data protection or privacy laws that regulate transborder data flows. By year 2014 the number of countries has increased over 100 [25].

A. *European Union*

European Union (EU) General Data Protection Regulation (GDPR) [3] defines rules for the protection of the fundamental rights and freedoms of natural persons with regard to the processing of personal data. Article 1 says that "The protection of natural persons in relation to the processing of personal data is a fundamental right." It sets limitation to where data can be stored and when it should not be kept anymore. General idea with GDPR and OECD recommendations [2] is that transborder data flows are acceptable, if all countries on the way have compatible legislation. Specifically GDPR Article 1(3) states: "The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data." GDPR recitals 101, 102, 103, 105, 107, 115, and 116 explain various cases for cross-border data transfers. On the other hand, governmental surveillance laws in some countries are in conflict with GDPR and make storing and processing personal data of EU citizens unacceptable in those countries.

B. *Australia and New Zealand*

Regulations in Australia and New Zealand make it difficult for enterprises to move sensitive information to cloud-providers that store data outside of Australian/New Zealand borders [26]. According to The Privacy Amendment Act [27] personal information can be transferred, but the Australian sender must take reasonable steps to ensure the recipient will comply with the Australian Privacy Principles (APPs). Even then the Australian sender remains liable for the recipient's behavior in this context.

C. *China*

Currently, in the Peoples Republic of China there is not a comprehensive data protection law, but various laws and regulations contain provisions that can be interpreted as citizen's right to privacy or reputation. In practice, The Decision on Strengthening Online Information Protection and National Standard of Information Security Technology – Guideline for Personal Information Protection within Information System for Public and Commercial Services are used as compliance references. In addition, provisions contained in other laws and regulations may be applicable depending on the industry or type of information [25].

According to the guideline, unless explicit consent from data subject, express authorization from laws or regulations or authorization from relevant authorities is acquired, personal information must not be transferred to a receiver outside the territory of the People's Republic of China. [25].

There are new laws being prepared: Personal Data Protection Law and Cybersecurity Law [25].

D. *Russia*

In Russia personal data localization requirements implemented by the Amendments to the Personal Data Law as of September 2015 mandate all personal data of Russian citizens to be stored in databases that reside in territory of the Russian Federation [28]. Personal data can still be duplicated to servers outside Russian borders, as long as other Russian laws regarding personal data are followed. The law does not restrict remote access to databases located in Russia. However, the laws are not quite exact and leave room for interpretation.

E. *USA*

In the United States of America (USA), there is no single, comprehensive federal (national) law regulating the collection and use of personal data [29]. However, there are many government policies and regulations that deal specifically with data privacy and residency issues. The Health Insurance Portability and Accountability Act (HIPAA) is a data privacy and security law designed to protect medical information. Payment Card Industry Data Security Standard (PCI DSS) is a set of policies to secure credit and debit cardholder information [30].

VI. Trustworthy Locating solutions

Several radio communication protocols can be used to implement proximity based locating. We will evaluate RFID, Bluetooth, ZigBee, Wi-Fi and cellular telecommunication systems. Locating can also be based on other than radio signal, for example, light and sound [10]. These are chosen because of their ubiquity and availability of reception technologies in server environments.

A. Proximity to RFID Reader

RFID (Radio Frequency IDentification) uses small tags, which each have a unique identification code. The code can be read remotely. There are passive and active tags. A passive tag does not have power source and when read, it gets necessary power burst from the reader via radio waves. Reading range of passive tags is from 10cm to 100cm, whereas active tags can be read from the distance of up to 500m [31].

RFID based systems are offered for real-time location needs. Asset tracking systems can be used in hospitals and offices to track equipment and paper files. Typically they are utilizing passive RFID tags which are readable from the distance of few meters.

Active RFID tags are more expensive than passive tags but still cost only few euros a piece. They need a power source, which may cause maintenance challenges in ordinary locating systems. Nevertheless, compared to other radio signal transmitters this is a reasonable price.

B. Bluetooth Networks

Bluetooth [32] was developed to serve as a personal area network (PAN). Bluetooth radio can reach devices from distance of less than 10m up to 100m, depending on the device class. Bluetooth specifies three classes with transmit power from 1mW to 100mW. Bluetooth Special Interest Group has announced that forthcoming Bluetooth 5 will have range up to 400m. Bluetooth 5 is expected to be published early 2017.

Bluetooth piconet can connect one master device with maximum seven active slave devices. In addition there can be 255 parked devices, which can be communicated with after being activated. This scheme can be extended as scatternet, where multiple piconets are linked by a shared device.

There are experiences that 100 devices can be detected using Bluetooth in 15 seconds [33]. More devices are possible if longer detection time is not a problem. Bluetooth standard does not set practical upper limit to the device count.

Recent Bluetooth transceiver component development has increased the range beyond 100m [34]. Also proposed mesh networking mode enables even longer distances for a Bluetooth network. The mesh network has a topology in which all devices can communicate with other devices either directly, if in range, or indirectly via one or more intermediate devices.

C. ZigBee Connectivity

Protocols intended for Internet of Things (IoT) networking, like ZigBee [35], resemble Bluetooth as far as our needs are concerned. IoT networks can be organized as router controlled star or as mesh kind of networks, similarly to Bluetooth. A ZigBee based solution utilizes mesh network, RSSI measurements and known locations of the static stations [36].

ZigBee device addresses are different from those of Bluetooth. Each ZigBee device has a unique eight byte EUI64 address [37]. When a device joins a ZigBee network, it sends an association request to the coordinator node, which has predefined node id. The ZigBee coordinator node assigns a 16-bit node id to the new member. Node ids can be used instead of the long EUI64 addresses. Node id based addressing allows 64000 devices in a single network. However, network coordinators can be linked to have several networks to work together.

D. Wi-Fi Networks

Wi-Fi is IEEE 802.11 based WLAN. Wi-Fi can use higher transmission power yielding longer geographical range than low energy Bluetooth and ZigBee. The range can be further extended with help of repeaters. The longer the range, the more inaccurate location information can be achieved by mapping all wirelessly connected devices to the location of the AP.

IEEE 802.11 has several variants being developed over years. The basic protocol can handle 2007 concurrent clients [38] by limiting the number range of the association identification (AID) field.

Wi-Fi Alliance has published Wi-Fi Aware specification that extends Wi-Fi capabilities to proximity detection [39]. It is designed for indoor alerting of potential customers in shopping centers and alike and should be by design capable of handling crowded environments.

E. Cellular Networks

With cellular communication systems the signal range is tens of kilometers. There are BTSs which are designed to serve small cells by using antennas that limit the coverage area.

Cellular networks mobile phone location capability has been proposed to locate data center servers [40]. The idea is that a server has a build in cellular phone with a SIM card and antenna installed outside the server room. Telecom operator can now regularly verify the location of the server.

VII. Discussion

Technical solutions should cope with the requirements explained in section III.A on page 3. Detailed analysis is subject to further work. Here we present some observations concerning the matter.

A. Common observations

Neither Bluetooth beacon, RFID readers, ZigBee static nodes nor Wi-Fi AP know their geographical coordinates. Locations could be stored to a database of the support system, but for our purpose the exact location of a BTS is not actually important. It is enough to know that a BTS is inside the data center premises and its signal range does not reach outside the premises.

All mentioned short range techniques operate on the same 2.4GHz band. They are susceptible to signal absorption, diffraction, interference and multipath propagation. At the moment we don't know if these radio technologies work reliably enough in noisy data center circumstances.

An essential question is that if we can locate a server dependably how to associate a physical server with a certain data record. It is actually the geolocation of data that we should be worried about, at least as much as location of compute servers.

However these technologies do have the advantage that relative distances and proximities are readily available along with local databases of known or seen devices, cf: Bluetooth pairing.

B. RFID

We could attach a passive tag inside every server and then use a hand held RFID reader to check, which servers are present in a data center. However, this would require considerable effort even in a modest data center because the reader should be brought next to every server, one by one. In addition, we would need a database to map physical server identifications, servers' identifications visible to software and geographical locations of the servers. Even then the results would not be trustable: somebody may replace some tags or will fail to maintain the database in case of server replacement or movements, etc.

Alternatively, we could install an active RFID tag to every server motherboard already at the factory and feed power from the server's power supply. With a battery backup the tags would serve logistic purposed already before being deployed for use, which alone can justify the cost of the tag. Data centers could have a reader collect signals from the tags [16]. A single reader in the middle of the data center, could cover area of up to 500m radius from a reader or even further [41].

RFID technology may have problems reading all tags reliably if the number of tags within reader's range is too high. This can occur in a data center with thousands of servers.

Further studies are required to clarify performance with large volume of active tags per reader. Also sufficient performance in noisy data center circumstances must be experimentally proven.

C. Bluetooth

Utilizing proximity to Bluetooth beacon, the resulting accuracy of location fits well with our need. Based on these assumptions, a large data center site may need several beacons.

Traditionally data center computers do not have Bluetooth transmitters build in. Nonetheless, adding Bluetooth capability to motherboard costs only about five euros.

The same uncertainties apply with Bluetooth as with active RFIDs: we should learn about the actual performance in realistic data center environment.

D. ZigBee

ZigBee is not intended to be used in data server environment. These can be problems with finding suitable chip sets to be integrated in to a server motherboard.

Again, same reservation apply concerning realistic performance of the technology as with Bluetooth and RFID.

E. Wi-Fi

Wi-Fi is on keen focus of retail shopping centers as an indoor location technology [42]. The motivation is to be able to push advertisements to potential customers walking by with a mobile phone in their hand.

Due to the maximum limit in addressing devices by a Wi-Fi AP, a large data center needs tens of APs, each of which should have a well-designed location to attract only about 1000 servers ("stations" (STA) in 802.11 parlance) in its range. This is a problem considering difficulties in estimating how radio signals propagate in a data center kind of environment.

When used for detecting servers there is no need get high bit rates through Wi-Fi and thus 1000 STAs should be possible throughput wise.

Wi-Fi chips are available in millions of portable or mobile devices. Adding Wi-Fi to a server motherboard should cost less than 5 euros, which is in the same price range as Bluetooth or RFID.

F. Cellular Telecommunication Systems

When we take into account possible blockage of radio signals from penetrating data center facilities, public cellular telecommunication system can be excluded from the alternatives because their antennas are outside the data center.

Even if we install a cellular BTS inside the data center, there is risk that the signal leaks to outside the premises. Because the technology itself is designed to work over tens of kilometers, this open possibility to forging location of a server to look like it is in the expected jurisdiction although it is not. This is an issue, if the data center is near border of a jurisdiction. Necessary hardware to add cellular network connectivity to every server is more costly than five euros, which is the price of the alternative technologies. Considering cost of BTSs or alternatively cost of arranging antennas outside the premises, total cost would be higher than with alternative technologies.

G. Internet Distances

Measurements based on distance bounding protocols can be used to calculate approximate location of any internet connected server on the Globe.

If the computer network is heterogeneous, delay based distance approximations will give misleading results. Calibration measurements can be used to create correction factors, but even then this method is not very accurate. Calibrations must be repeated whenever network topology changes. There can also be fast “dark fiber” connections between data centers of a provider. Those connections are not visible to the Internet. Also proxy servers may distort delay measurements as well as temporary fluctuations of workloads and network congestion.

H. *Software Aspects*

Each of the presented technologies require special locating support software to be added to the servers. For sufficient dependability, this software should be included as trusted modules, using, for example, Trusted Platform Module (TPM) as root of trust. To make cheating more difficult, the locating software should be carefully tested and reviewed against vulnerabilities. It should be embedded at a low software layer, like part of Unified Extensible Firmware Interface (UEFI), so that it cannot be easily tampered in the target computer.

Indeed current TPM attestation technology already includes support for “trusted location” albeit in terms of a cryptographic hash without specific location semantics. Generation of the data to be placed in a TPM relating to location still however has to be generated reliably.

One issue here is that computation of proximities and distances over hashes of this kind is not possible; technologies such as homomorphic encryption may provide a solution here [43].

I. *Hardware Aspects*

Radio based solutions need transceiver chips that are not part of a typical server hardware. Further studies and experiments are needed to identify the most feasible technical solutions.

VIII. Conclusion

Legal restrictions to geographical location of data processing are driven by governments willing to protect their citizens and enterprises against data breaches, which may become easier if data is stored or transmitted at or via locations that do not have as strict data protection legislation as the domestic one. The reverse can also apply: a government may have easier access to domestic databases.

Another motivator is worry that confidentiality of information can become compromised. Such information includes industrial secrets, health records, and private personal data.

All mentioned techniques have challenges when being applied to locating servers in to a data center site.

A further interesting challenge will be to utilize data from multiple geolocating sources and understand how this all together can be unified to provide a more robust and fault-tolerant geolocation mechanism.

This paper focuses on the problem domain. Further work will focus on cheating and attack patterns and explore the solution domain, including locating data records.

Acknowledgment

The authors would like to thank Professor Tuomas Aura of Aalto University. This work was made under the Finnish Dimecc Cyber Trust Program.

References

[1] N. Palad and A. Michalas, "’One of our hosts in another country’: Challenges of data geolocation in cloud storage," in *4th International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE)*, 2014.

[2] *https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf,* OECD Publishing, 2013.

[3] *Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation),* 2016.

[4] M. Bartock, M. Souppaya, R. Yeluri, U. Shetty, J. Greene, S. Orrin, H. Prafullchandra, J. McLeese, J. Mills, D. Carayiannis, T. Williams and K. Scarfone, "Trusted Geolocation in the Cloud: Proof of Concept Implementation," National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, MD, USA, 2015.

[5] "Data center," Wikipedia, 22 Sep 2016. [Online]. Available: https://en.wikipedia.org/wiki/Data_center. [Accessed 26 Sep 2016].

[6] P. Mell and T. Grance, *The NIST Definition of Cloud Computing,* National Institute of Standards and Technology, U.S. Department of Commerce, 2011.

[7] "Cloud computing," Wikipedia, 23 Sep 2016. [Online]. Available: https://en.wikipedia.org/wiki/Cloud_computing. [Accessed 27 09 2016].

[8] "Modular data center," Wikipedia, 25 Feb 2016. [Online]. Available: https://en.wikipedia.org/wiki/Modular_data_center. [Accessed 29 Sep 2016].

[9] H. Liu, H. Darabi, P. Banerjee and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews,* vol. 37, no. 6, pp. 1067-1080, Nov 2007.

[10] R. Mautz, "Indoor Positioning Technologies," Feb 2012. [Online]. Available: http://e-collection.library.ethz.ch/eserv/eth:5659/eth-5659-01.pdf. [Accessed 28 Sep 2016].

[11] A.-M. Roxin, J. Gaber, M. Wack and A. Nait-Sidi-Moh, "Survey of Wireless Geolocation Techniques," *2007 IEEE Globecom Workshops,* pp. 1-9, 2007.

[12] "Timing advance," Wikipedia, 16 Jan 2016. [Online]. Available: https://en.wikipedia.org/wiki/Timing_advance. [Accessed 27 Sep 2016].

[13] A. A. Albeshri, C. Boyd and J. Gonzalez Nieto, "Geoproof : proofs of geographic location for cloud computing environment," in *Proceedings of the 32nd International Conference on Distributed Computing Systems Workshops 2012*, Macau, China, 2012.

[14] A. Günther and C. Hoene, "Measuring Round Trip Times to Determine the Distance between WLAN Nodes," in *NETWORKING'05 Proceedings of the 4th IFIP-TC6 international conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communication Systems*, 2005.

[15] K. Benson, R. Dowsley and H. Shacham, "Do you know where your cloud files are?," in *Proceedings of the 3rd ACM workshop on Cloud computing security workshop, CCSW '11*, New York, NY, USA, 2011.

[16] K. Jayaram, D. Safford, U. Sharma, V. Naik, D. Pendarakis and S. Tao, "Trustworthy geographically fenced hybrid clouds," in *Proceedings of the 15th international middleware conference (2014)*, 2014.

[17] T. Böttger, F. Cuadrado, G. Tyson, I. Castro and S. Uhlig, "Open Connect Everywhere: A Glimpse at the Internet Ecosystem through the Lens of the Netflix CDN," Queen Mary University of London, 17 Jun 2016. [Online]. Available: http://arxiv.org/abs/1606.05519.

[18] S. Itzhak, "Global Positioning Using Planetary Constants". U.S.A. Patent 7,822,549, 26 Oct 2010.

[19] R. Artes and D. Bonte:, "Goods Tracking Technologies," ABIresearch, Oyster Bay, New York, USA, 2016.

[20] S. J. Vaughan-Nichols, "How Google--and everyone else--gets Wi-Fi location data," ZDNet, 16 Nov 2011. [Online]. Available: http://www.zdnet.com/article/how-google-and-everyone-else-gets-wi-fi-location-data/. [Accessed 16 Sep 2016].

[21] WhatIs, "Definition: data sovereignty," TechTarget, Mar 2013. [Online]. Available: http://whatis.techtarget.com/definition/data-sovereignty. [Accessed 18 Sep 2016].

[22] WhatIs, "Definition: data residency," TechTarget, Jun 2015. [Online]. Available: http://searchcloudcomputing.techtarget.com/definition/data-residency. [Accessed 23 Sep 2016].

[23] L. Determann, E. Bekeschenko and V. Perevalov, "Residency Requirements for Data in Clouds—What Now?," 2015. [Online]. Available: http://www.globalequityequation.com/files/Uploads/Documents/Equity%20Equation/Residency%20Requirements%20for%20Data%20in%20Clouds%20--%20What%20Now.pdf. [Accessed Sep 2016].

[24] C. Kuner, "Regulation of Transborder Data Flows under Data Protection and Privacy Law: Past, Present and Future," OECD Publishing, 2011.

[25] DLA Piper, "DLA Piper's Data Protection Laws of the World Handbook," 2016. [Online]. Available: https://www.dlapiperdataprotection.com/. [Accessed Sep 2016].

[26] "Australia Data Privacy Laws," Symantec Blue Coat, 2016. [Online]. Available: https://www.bluecoat.com/resources/cloud-governance-data-residency-sovereignty/australia-data-privacy-laws. [Accessed 22 Sep 2016].

[27] *Privacy Amendment (Enhancing Privacy Protection) Act,* Australia, 2012.

[28] A. L. Gallia, L. P. McLoughlin, A. S. Khaskelis and M. A. Voltchenko, "Russian Federation: Russia's Personal Data Localization Law Goes Into Effect," Duane Morris LLP, Mondaq, 15 Oct 2015. [Online]. Available: http://www.mondaq.com/russianfederation/x/435890/Data+Protection+Privacy/Russias+Personal+Data+Localization+Law+Goes+Into+Effect.

[29] I. Jolly, "Data protection in United States: overview," Practical Law, 1 Jul 2016. [Online]. Available: http://uk.practicallaw.com/6-502-0467. [Accessed 14 Sep 2016].

[30] M. Rouse, "Definition: data residency," TechTarget, Jun 2015. [Online]. Available: http://searchcloudcomputing.techtarget.com/definition/data-residency.

[31] M. Roberti, "What Is an RFID Reader's Maximum Range?," *RFID Journal,* 19 Jan 2014.

[32] *Specification of the Bluetooth System, Version 4.2,* Bluetooth SIG, 2014.

[33] "How many active Bluetooth devices can I reliably detect in a single space?," Stackexchange, Electrical Engineering, 2011. [Online]. Available: http://electronics.stackexchange.com/questions/21991/how-many-active-bluetooth-devices-can-i-reliably-detect-in-a-single-space. [Accessed 14 Sep 2016].

[34] "Cypress Unveils a Bluetooth Low Energy Module with 400-Meter Range and a Module with Bluetooth 4.2 Support," Cypress Semiconductor, San Jose, USA, 15 Mar 2016. [Online]. Available: http://www.cypress.com/news/cypress-unveils-bluetooth-low-energy-module-400-meter-range-and-module-bluetooth-42-support. [Accessed 9 Sep 2016].

[35] *ZigBee 3.0 standards,* ZigBee Alliance, 2014.

[36] O. Hernandez, V. Jain, S. Chakravarty and P. Bhargava, "Position Location Monitoring Using IEEE 802.15.4/ZigBee® technology," NXP Semiconductors Netherlands B.V., 2014. [Online]. Available: https://www.nxp.com/files/microcontrollers/doc/brochure/PositionLocationMonitoring.pdf. [Accessed 6 September 2016].

[37] "Maximum number of ZigBee devices," Stackexchange, Stackoverflow, 2013. [Online]. Available: http://stackoverflow.com/questions/20175647/maximum-number-of-zigbee-devices. [Accessed 14 Sep 2016].

[38] IEEE-SA Standards Board, *IEEE Std 802.11-2007, IEEE Standard for Information Technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY,* 2011.

[39] "Wi-Fi Aware," Wi-Fi Alliance, 2015. [Online]. Available: http://www.wi-fi.org/discover-wi-fi/wi-fi-aware. [Accessed 29 Sep 2016].

[40] D.-Y. Yu, A. Ranganathan, R. J. Masti, C. Soriente and S. Capkun, "SALVE: Server Authentication with Location VErification," in *22nd ACM International Conference on Mobile Computing and Networking (MobiCom 2016)*, 2016.

[41] C. Swedberg, "Iotera Develops Active RFID Tag With 4-Mile Read Range," *RFID Journal,* 14 Jan 2014.

[42] P. Connolly and D. Bonte:, "Wi-Fi Indoor Location Applications and Revenues," ABIresearch, Oyster Bay, New York, USA, 2016.

[43] M. Ekholm, "Applications of Homomorphic Encryption," 29 Sep 2015. [Online]. Available: https://aaltodoc.aalto.fi/handle/123456789/18158. [Accessed 29 Sep 2016].

[44] E. Schmieders, A. Metzger and K. Pohl, "A Runtime Model Approach for Data Geo-location Checks of Cloud Services," in *Service-Oriented Computing, 12th International Conference, ICSOC 2014*, Paris, France, 2014.

[45] Y. Chou, "Cloud Computing for IT Pros (2/6): What Is Cloud," Microsoft Inc., 17 Dec 2010. [Online]. Available: https://blogs.technet.microsoft.com/yungchou/2010/12/17/cloud-computing-for-it-pros-26-what-is-cloud/. [Accessed 22 Sep 2016].