EasyChair Preprint

№ 820

Resilience and Security in Computer Science

Loredana Vigliano

March 12, 2019

RESILIENCE AND SECURITY IN COMPUTER SCIENCE

Loredana Vigliano
University of Rome Tor Vergata, Italy, vigliano@mat.uniroma2.it

ABSTRACT. The computer resilience is the ability of a system to guarantee services despite adverse agents, progressively more difficult to treat, such as: system wear, software upgrades, system failures, security attacks. This goal can be achieved by several techniques, and many companies offer services to allow business continuity.
The study and the analysis to prevent disasters or security attacks becomes thus as study and analysis of system and network anomalies. Therefore, understanding the behavior of data flowing over the network allows us to deduct rules or laws that underlie to this structure, increasing our knowledge about social data and hidden social behaviors. Thus, the big data analysis helps us predicting the future evolution of all things.

Keywords: business continuity,resilient company, fake identities, hacker, big data, data science, artificial intelligence, machine learning, predictive algorithms.

INTRODUCTION

In computing science, resilience is the capability of a system to adapt to the conditions of use and to ensure long term availability of the services offered.

Resilience of computer systems (or index of fragility[1]) is therefore the ability of such a system to adapt to ensure the agreed services, or more generally the needed flexibility to assume new behaviors if the earlier are no longer available or suitable.

The arising of big data, expecially those of social networks and search engines, changes the resilience meaning : resilience is the ability of a server, network, storage system, or an entire data center, to recover quickly and continue operating even when there has been an equipment failure, power outage or other disruption. [15]

These goals can be achieved using different techniques depending on the type of "trauma" supported by the system. A very general classification of the types of defense to take is closely linked to the types of issues that can arise, and the issues can be summarized as follows [2] (although often the problem's boundaries are not so clear and faults and/or attacks may be chained).

DIFFERENT TYPES OF PROBLEMS :

Extensive Use Problems or Hardware Failures

These problems can be solved by replacement and redundancy techniques. Take the case where you have, for example, a floppy disk HD at the end of its working life and therefore with high risk of failure. The resilience can be obtained by duplicating the resource available for the writing of data. For example you can write data on multiple disks to safeguard information against the failure of a single disk.

Therefore the solution can be a redundancy mechanism for critical parts. But it may be also the distribution of an application on multiple computers so to absorb the workloads.

One of easiest techniques is **Mirroring**[10] and allows you to have in the storage unit two copies of the disc (or just some pre-selected important archives); when an I/O finds an error, processing is not stopped because it can use the alternative copy.

A second level of fault tolerance is achieved by the technique of **duplexing**, which consists of a duplication of the disks' control unit (controller) instead of discs.

The third level covers **duplication of the whole system**,[3] for example a server of a local network or a mainframe in case of a large system.

There are other cheaper techniques available for the partial duplication of the system, that are denoted with the acronym RAID (Redundant Array of Independent Disks)[4]. This technology distributes data across a group of disks, so that it is possible to mathematically rebuild any data eventually lost from one disk.

Fault Tolerance and Clustering Techniques.

A more complex system, always relative to power supply of the active devices, is the replication of the power supply unit; if the main power supply should be damaged, the system will still work thanks to one or more redundant power supplies. The fault tolerance of course equals the number of redundant power supplies used in the system: to put it simply, if a system has three power supplies and all of them break down at once, the system stops.

Data center resilience is often achieved through the use of redundant components, subsystems, systems or facilities..Techniques, such as server clustering, support redundant workloads on multiple physical servers. When one server in the cluster fails, another node takes over with its redundant workloads. [15]

A case study : Google.

Any professional system uses techniques like this to save data and ensure their reliability thus maintaining confidence of users that rely on those services. A company like Google (specializing in Internet-related services and products that include online advertising technologies, search, cloud computing, software, and hardware), for example, has proprietary techniques for data recovery based on Bigtable. Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers. [16].

Bigtable relies on a highly-available and persistent distributed lock service called Chubby . A Chubby service consists of five active replicas, one of which is elected to be the master and actively serve requests. The service is alive when a majority of the replicas are running and can communicate with each other. [16].

In addition Google replicates the user data between 3 and 5 times on different servers, in order to ensure the reliability, but also, as we will see, not to lose important information useful for data analysis[5]. Indeed, Google's mission statement from the outset was "to organize the world's information and make it universally accessible and useful,".[18].

Pirate Attacks and hackers (Mainly Software)[6][7]

Imagine someone trying to force open the door of your home. Now imagine that occurs 60,000 times a day.

This is the number of times a medium size IT company's infrastructure is under attack. In the US only, over the past five years, there have been 354 million privacy violations. In January 2009, a single episode of cyber security has compromised 130 million credit cards[1][8].

No technology or tool is without risk. Wherever people gather, there are bound to be criminal elements on the fringe of the crowd. The Internet is no different. Almost daily it seems we hear about a new virus spreading through millions of computers; or about companies and government agencies losing sensitive data of employees, customers, and citizens.

As complex as the technology used to create and develop the Internet is, so too is the network of online criminals and their cyber arsenal of viruses, Trojans and other forms of malware used to dupe unsuspecting consumers and even steal their identities. Internet threats are a clear and present danger to

society, as the potential economic rewards for criminals are enormous and the curtain of anonymity behind which they can hide is equally heavy. Security threats such as viruses, worms, Trojans and spyware are often designed to exploit vulnerabilities in common software products forcing software developers to constantly develop patches and other fixes to keep emerging malware at bay. [19].

A distinguished description can be made for "hacker" role with its different characterizations.
In computing, a **hacker** is any highly skilled computer expert capable of breaking into computer systems and networks using bugs and exploits. Depending on the field of computing it has slightly different meanings, and in some contexts has controversial moral and ethical connotations [21]:

.

- Black hats are hackers with malicious intentions, and steal, exploit, and sell data. They are usually motivated by personal gain.
- Grey hats are hackers who are neither good nor bad, and often include people who hack 'for fun' or to 'troll'. They may both fix and exploit, though grey hats are usually associated with black hat hackers.
- Black hats are hackers with malicious intentions, and steal, exploit, and sell data. They are usually motivated by personal gain.

Moreover a cracker (also known as a black hat hacker) is an individual with extensive computer knowledge whose purpose is to breach or bypass internet security or gain access to software without paying royalties. The general view is that, while hackers build things, crackers break things. [20].

More Generally

Unfortunately, companies face other types of risks as well as fraudulent attacks. There are business risks, such as auditing, new product sales, future marketing efforts, standards compliance. Finally, risks due to natural disasters, blackouts, acts of war or economic crises. All of these risks can be confined if you identify weaknesses and build a plan accordingly.

ENVIRONMENTS (AND STATISTICS)

Computer resilience environments are those related to business continuity and disaster recovery.

Business continuity is defined as the ability of the company to continue its business even during or after adverse events that can hit it. [1] The business continuity planning and service is called business continuity plan (BCP) ([2]) and commonly consists of a comprehensive process that identifies potential damages that threaten the organization, providing also a framework that allows you to increase the resilience and responsiveness in order to safeguard the interests of the stakeholders, the production activities, the image, reducing the risks and consequences in terms of management, administrative, legal.
Disaster recovery, in computer science and especially in computer security, is the set of technological measures and logistics / organizational measures to restore systems, data and infrastructure necessary to the provision of business services for companies, associations or entities when in serious emergencies that involve the regular activities.

The impact of these emergencies is such that it is estimated that most of the large companies spend between 2% and 4% of their IT budget on disaster recovery management planning in order to avoid further losses in case activities should be stopped due to the loss of data and IT infrastructure. Companies that have suffered disasters with heavy loss of data, about 43% no longer resumed activity, 51% had closed within two years, and only 6% has managed to survive in the long term. [1][2] The disasters computer with large losses of data in most cases can cause the failure of the enterprise or organization, so investment in appropriate recovery strategies becomes an obvious choice.

COMPANIES FOR RESILIENCE

All these issues generate an important business for companies specialised in disaster recovery and high reliability services such IBM that relies on this sector a large part of its revenues, with their declared 312 resiliency centers in 68 countries

Here are the most significant phrases used for advertising of this important business just to emphasize how much this problem is important felt so in all types of company and how, currently, the resilience is very critical for our daily life:

Are you ready?
Where's vulnerable your company? Your risk management activities allow you to respond quickly to daily threats? Do they help you better manage your business risks?
Find out if your company is ready to face the risks. The self-assessment tool for risk management helps you understand how, throughout your company, the three key disciplines of a mature risk management are implemented:

effective control risk
robust IT infrastructures
a culture of risk awareness

It is no longer sufficient to develop an expensive infrastructure and have a set of tools to minimize the impact of risk and return to normal activities after a disaster. Security cannot be solely the responsibility of people involved in the definition of business rules or those of a specific business unit. Modern companies have to develop a business strategy that takes into account also intelligent risk management. Companies, cities, communities, government agencies and civil society share-and influence-critical systems of our planet.[1]

In this advertising we can note some keywords like 'disaster', 'business', 'planet', etc. , carefully choosed to underline the importance of the resilience.
Resiliency services offered are analytics tools that can leverage data sources in new ways to enhance the effectiveness of threat analysis, fraud detection, etc. and execution of predefined policies using authorative sources to attest the identities, roles, attribute or ather context.

ADDITIONAL ISSUES NETWORK RELATED.

In connection with previously topics, additional issues raise when considering network connections (we have already seen in the section "Pirate Attacks and hackers"). Systems should not be considered as stand alone objects but as participating to a world wide interconnected network, where they are continuously querying and responding to search for information and share of results. This new paradigm spreads a single issue to all systems in a network (e.g. failure of network nodes holding distributed data or incomplete process workflows on distributed servers).

In the world of 'democratized' Public Clouds, node failures are not just probable, they are expected. This requires database technology that can withstand failure and continue to perform.
Thus, new and alarming scenarios can be imagined, and since they are especially dangerous for archived data and information, it's important to preview them and be prepared to manage them if they eventually happen.
The presence of great attention to these issues is demostrated by the following case study from Netflix.

A case of resilient company : Netflix

Netflix Inc. is an American multinational entertainment company founded on August 29, 1997, in Scotts Valley, California, by Reed Hastings and Marc Randolph. It specializes in and provides streaming media and video on demand online and DVD by mail. In 2013, Netflix added film and television production, as well as online distribution. In October 2016, Netflix reported over 86 million subscribers worldwide, including more than 47 million in the United States [17].

The following case study is published on their public blog :

On Sept 25th, 2014 AWS notified users about an EC2 Maintenance where "a timely security and operational update" needed to be performed that required rebooting a large number of instances. (around 10%) On Oct 1st, 2014 AWS sent an updated about the status of the reboot and XSA-108.
While we'd love to claim that we weren't concerned at all given our resilience strategy, the reality was that we were on high alert given the potential of impact to our services. We discussed different options, weighed the risks and monitored our services closely. We observed that our systems handled the reboots extremely well with the resilience measures we had in place. These types of unforeseen events reinforce regular, controlled chaos and continued to invest in chaos engineering is necessary. In fact, Chaos Monkey was mentioned as a best practice in the latest EC2 Maintenance update.
Our commitment to induced chaos testing helps drive resilience, but it definitely isn't trivial or easy; especially in the case of stateful systems like Cassandra. The Cloud Database Engineering team at Netflix rose to the challenge to embrace chaos and runs chaos monkey live in production last year. The number of nodes rebooted served as true battle testing for the resilience design measures created to operate cassandra.[14]

"*When we got the news about the emergency EC2 reboots, our jaws dropped. When we got the list of how many Cassandra nodes would be affected, I felt ill. Then I remembered all the Chaos Monkey exercises we've gone through. My reaction was, "Bring it on!".*" - Christos Kalantzis - Engineering Manager, Cloud Database Engineering. [14].

And finally :

Repeatedly and regularly exercising failure, even in the persistence layer, should be part of every company's resilience planning. [14].

Data center resilience must be a planned part of a facility's architecture and is usually associated with other disaster planning and data center disaster-recovery considerations such as data protection. The adjective *resilient* means "having the ability to spring back." [15].

RESILIENT DEVELOPMENT

We can now give a wider **definition** of Computer Resilience, or even better of the development process that is involved by this issue:

Resilient Development allows to innovate the business strategy towards sustainable development, and to prevent and manage stress and crisis.

Therefore it is important to have tools to anticipate and prevent events before they happen, especially with the help of the study of unusual activity of data flow over the network. This is particularly valuable in relation to what we will explain in the next chapter.

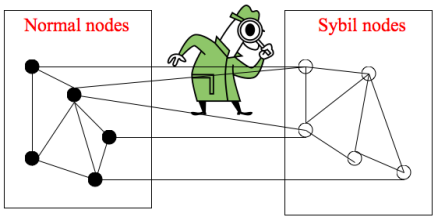
FAKE IDENTITIES – SYBIL ATTACK – SOCIAL DEFENSE

Social networks have attracted worldwide attention because of their potential to address millions of users and possible future customers. The potential of social networks is often misused by malicious users who extract sensitive private information of unaware users. Then. another kind of problem that makes network and systems vulnerable is that of Fake Identities, i.e. those identities that appear for example in social networks and ask the "friendship" with the purpose to steel money, falsify voting or otherwise acquire private information for cheating. One of the most common ways of performing a large-scale data harvesting attack is therefore the use of fake profiles, where malicious users present themselves in profiles impersonating fictitious or real persons.[22] Sometimes they are even used to create an artificial audience to review products or to make a business grow by making it popular through many profiles, or even to spread opinions and ideologies. According to Facebook, 5% to 6% of registered Facebook accounts are fake accounts [22] and it is estimated that we grant friendship with false identities with an average probability of 30%[6].

In general, a social network can also be modeled as a graph. Hence, we define $G = (V, E)$, where V is a set of n nodes. Each participant of a social network, namely a user account, is represented by a node. E

is a set of edges. An edge connects two nodes. If two nodes are adjacent, this means that there is a social connection established between the users modeled by the nodes (e.g. friendship)[22].

Figure 1. Fake Identities



It's almost impossible to identify automatically these fake identities, but, assuming that the false accounts have social data and the set of these accounts has the structure of the nodes of a subnet social, the study and analysis of this structure will be able to find the differences between a real social subnet or a fake one[7]. Generally profiles that do not display social activities and a high number of friends are more likely to be perceived as fake than profiles that display social activities and interactions with others[22].

Similar issue is a Sybil Attack. Named after the case study of a woman with multiple personality disorder, a Sybil attack is a type of security threat when a node in a network claims multiple identities[23].
Most networks, like a peer-to-peer network, rely on assumptions of identity, where each computer represents one identity. A Sybil attack happens when an insecure computer is hijacked to claim multiple identities. Problems arise when a reputation system (such as a file-sharing reputation on a torrent network) is tricked into thinking that an attacking computer has a disproportionally large influence. Similarly, an attacker with many identities can use them to act maliciously, by either stealing information or disrupting communication. [23].
Sybil prevention techniques based on the connectivity characteristics of social graphs can also limit the extent of damage that can be caused by a given Sybil attacker while preserving anonymity, though these techniques cannot prevent Sybil attacks entirely, and may be vulnerable to widespread small-scale Sybil attacks.[24].

The study and analysis of the data flow anomalies on web social graphs and network traffic is the main feature of Social Defense. Understanding the behavior of data flowing over the network allows us to deduct rules or laws that underlie to this structure and, as a result, any anomaly or deviation from these rules can be identified as a possible attack to computer security.
To achieve the purpose described above, even in Italy group has been created that is responsible to analyze the Italian Internet traffic to understand how our network is vulnerable to attack (Cyber Attack). The group works by analyzing large amount of data that travels over the network and using Data Mining techniques to find, on interchanges, tiny streams of data that have anomaly and could therefore be an incursion[7][8].

DATA SCIENCE AND BIG DATA FOR PREDICTIVE USE

"Data is the new oil, the new natural resource of our time" [25]. Data is increasingly cheap and ubiquitous. We are now digitizing analog content that was created over centuries and collecting myriad new types of data from web logs, mobile devices, sensors, instruments, and transactions. IBM estimates that 90 percent of the data in the world today has been created in the past two years. [26].
At the same time, new technologies are emerging to organize and make sense of this avalanche of data. We can now identify patterns and regularities in data of all sorts that allow us to advance scholarship, improve the human condition, and create commercial and social value. The rise of "big data" has the potential to deepen our understanding of phenomena ranging from physical and biological systems to human social and economic behavior.[26].

As said understanding the behavior of data flowing over the network allows us to deduct rules or laws that underlie to this structure and, as a result, any anomaly or deviation from these rules can be identified as a possible attack to computer security.

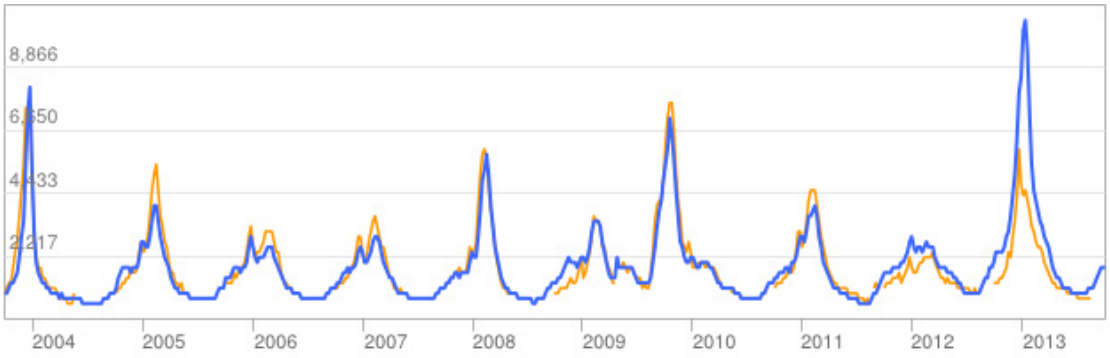
The importance of finding the laws that describe the social behavior of great multitudes of people, is not only useful for what was said up to now, but it is of top importance for the knowledge itself[9].
In fact, the great challenge of Data Science, studying big data available on the Internet, is to learn about human behavior and, through the digital traces that they leave, predict the future evolution of all things (or almost). Ambitious, but all in all practical and proved by some realistic solutions[10][11].

As an example we report a project of Google, Google Flu Trends ([12]), which, through the mapping of keywords searched on his engine in times of flu epidemic, has managed to display geographically (geolocalize) and to quantify the progress in time and in the world of a certain type of flu that developed in the East and then in Western countries. This mapping was foreseeing, quite precisely, the path of the flu in the world, the highest peak of the disease in different countries, etc.[12].. All this long before (about a month) that the American Health Institute published the same figures but based on established reports, not prediction.

A case of predictive project : Google Flu Trends

In November, 2008, Google launched Google Flu Trends (GFT), an internet-based surveillance tool that uses aggregated Google search data to estimate influenza activity in near-real time [28][30]. To account for evolving online search behavior for health information, GFT models are updated annually using the most recent official surveillance data, where available, and to utilize any newly developed modeling techniques. In the United States, GFT models outpatient influenza-like illness (ILI) using publicly available ILI surveillance data provided by the U.S. Centers for Disease Control and Prevention (CDC). The CDC's sentinel provider surveillance system, known as ILINet, is a collaborative effort between the CDC, state and local health departments, and health care providers that estimates weekly the proportion of health care provider visits that are due to ILI [29]. [30].
During the five years of data on which the original GFT model for the United States was built and tested, only seasonal influenza outbreaks occurred. [30].
Estimates of ILI produced by the GFT model developed in 2008 correlated highly with historical CDC ILI data, and GFT has since expanded to include 28 countries and 39 languages. [30].
The idea behind Google Flu Trends (GFT) is that, by monitoring millions of users' health tracking behaviors online, the large number of Google search queries gathered can be analyzed to reveal if there is the presence of flu-like illness in a population. Google Flu Trends compared these findings to a historic baseline level of influenza activity for its corresponding region and then reports the activity level as either minimal, low, moderate, high, or intense.

Figure 2. Google Flu Trends
Annual U.S. Flu Activity – Mid-Atlantic Region
(blue: Google Flu Trends, yellow: CDC Data
Data Source : Google Flu Trends Project
Credit : cc by 2.0 www.flickr.com)



But then, GFT failed—and failed spectacularly—missing at the peak of the 2013 flu season by 140 percent. When Google quietly euthanized the program, it turned the poster child of big data into the poster child of the foibles of big data.[31].
But GFT's failure doesn't erase the value of big data. What it does do is highlight a number of problematic practices in its use. [31].

Nature reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States. This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data , what lessons can we draw from this error?.[27].
The problems is not limited to GFT.

Although these studies have shown the value of these data, we are far from a place where they can supplant more traditional methods or theories . [27].

"Big data hubris" is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis. Elsewhere, there are enormous scientific possibilities in big data . However, quantity of data does not mean that one can ignore foundational issues of measurement and construct validity and reliability or resilience and dependencies among data. The core challenge is that most big data that have received popular attention are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis. [27].
The initial version of GFT was a particularly problematic marriage of big and small data. Essentially, the methodology was to find the best matches among 50 million search terms to fit 1152 data points. [27].

Google Flu Trends is now no longer publishing current estimates. Historical estimates are still available for download, and current data are offered for declared research purposes[32]. Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data. [27]. By today, similar projects such as the flu-prediction project[33] by the institute of Cognitive Science Osnabrück carry the basic idea forward, by combining social media data e.g. Twitter with CDC data, and structural models that infer the spatial and temporal spreading [34] of the disease.

The value of the data held by entities like Google is almost limitless, if used correctly. That means the corporate giants holding these data have a responsibility to use it in the public's best interest. [31].

ARTIFICIAL INTELLIGENCE , MACHINE LEARNING AND PREDICTIVE ALGORITHMS FOR RESILIENCE

Based on what have been said, it should be clear that resilience development in computer science involves much more scientific disciplines than expected at beginning. Among these disciplines we find: Artificial Intelligence, predictive algorithms, Data Mining, Machine learning and all founding sciences like mathematics, statistics, linear algebra, probability theory and as well as the studies on algorithms and calculus complexity.

Artificial Intelligence

It's very difficult to give an unique definition for Artificial Intelligence, here anyone there are :

> "The study of mental faculties through the use of computational models" (Charniak and McDermott, 1985).
> "A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes" (Schalkoff, 1990).
> "The art of creating machines that perform functions that require intelligence when per- formed by people" (Kurzweil, 1990).
> "The branch of computer science that is concerned with the automation of intelligent behavior" (Luger and Stubblefield, 1993). [35].

In computer science, an ideal "intelligent" machine is a flexible rational agent that perceives its environment and takes actions that maximize its chance of success at some goal.[36] Colloquially, the

term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving".[36]
These processes are closely linked to the notion of resilience.
Artificial Intelligence technologies involves natural language processing, information extraction, conceptual knowledge engineering, applied ontologies, knowledge-based systems, linguistic resources production, linguistic agents and semantic web.

Data protection and the ability to find alternative solutions and paths on the network graph and on systems that use it, resemble more and more the behavior and brain neurons ability to solve problems in the network of our brain, in case of damage. The study of Artificial Intelligence just develops this theme, and the one of learning and knowledge representation, with the help of many scientific disciplines such as mathematics, statistics, robotics, probability calculations, etc., not only on a single computer but on the global network [13].

Predictive Algorithms

An important characteristic of an intelligent agent is its ability to learn from previous experience in order to predict future events. The mechanization of the learning process by computer algorithms has led to vast amounts of research in the construction of predictive algorithms. Predictive algorithms can play a crucial role in systems management. The ability to predict service problems in computer networks, and to respond to those warnings by applying corrective actions, brings multiple benefits. First, detecting system failures on a few servers can prevent the spread of those failures to the entire network. For example, low response time on a server may gradually escalate to technical difficulties on all nodes attempting to com-municate with that server. Second, prediction can be used to ensure continuous provision of network services through the automatic implementation of corrective actions. For example, prediction of high CPU demand on a server can initiate a process to balance the CPU load by rerouting new demands to a back-up server.

Prediction techniques. Once the problem is well characterized, there are often a wide variety of prediction techniques available. In some cases they rely on classical time-series analysis, whereas in other cases they employ data-mining techniques. An important factor that differentiates among techniques is whether or not the model is homogeneous through time. [37].
Then Predictive algorithms play a crucial role in resilient systems management.

Machine learning

Machine learning provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data.[38].
Machine learning studies computer algorithms for learning to do stuff. We might, for instance, be interested in learning to complete a task, or to make accurate predictions, or to behave intelligently. The learning that is being done is always based on some sort of observations or data, such as examples, direct experience, or instruction. So in general, machine learning is about learning to do better in the future based on what was experienced in the past.
The emphasis of machine learning is on automatic methods. In other words, the goal is to devise learning algorithms that learn automatically without human intervention or assistance. The machine learning paradigm can be viewed as “programming by example.” Often we have a specific task in mind, such as spam filtering. But rather than program the computer to solve the task directly, in machine learning, we seek methods by which the computer will come up with its own program based on examples that we provide.
Machine learning is a core subarea of artificial intelligence. It is very unlikely that we will be able to build any kind of intelligent system capable of any of the facilities that we associate with intelligence, such as language or vision, without using learning to get there. These tasks are otherwise simply too difficult to solve. Further, we would not consider a system to be truly intelligent (and resilient) if it were incapable of learning since learning is at the core of intelligence.
Although a subarea of AI, machine learning also intersects broadly with other fields, especially statistics, but also mathematics, physics, theoretical computer science and more.[39].

The collection of all these types of studies, with its multidisciplinary nature, it is therefore extremely useful for our social life and applies to endless fields: health, marketing, finance, cognitive and

behavioral aspects and, as it has seen, to support the computer resilience, seen as safeguarding of knowledge, prevention of criticality and security privacy.

CONCLUSIONS

In this article we explored different definitions of computer resilience, going step by step more in depth.
First looking at organization and flexibility needed by a system to adapt and provide continuous availability despite of hardware or software failures or external attacks to the system itself, describing different techniques to achieve these goals.
Presented case study: Google.

Next computer resilience definition gets stronger and predictive in computer networks and network related issues: failure of nodes where important distributed information are stored, external attacks to violate security, privacy and information steeling , for example on social networks (fake-identities, Sybil attack).
Presented case study : Netflix.
An overview on important business for companies specialised in commercial resilience was also presented.

Another topic covered is the importance and reasons for safety of these data, as a safeguard for knowledge, but also, through the digital traces that we leave, as predicting the future evolution of all things.
Thus is in fact, the great challenge of Data Science, studying big data available on the Internet, to learn human behavior, illness trends, cellular and biological development, economic and financial progress, performance of the Stock Exchange, marketing overview, and so on.
Presented case study : Google Flu Trends (with its faults).

The multidisciplinary nature of these studies will widen more and more the idea of resilience.
The collection of all these types of studies it is therefore extremely useful for our social life and applies to endless fields: health, marketing, finance, cognitive and behavioral aspects and, as it has seen, it applies to support the computer resilience, intended as safeguarding of knowledge and prevention of criticality and security privacy.

REFERENCES

.[1] IBM Global CIO Study http://www-05.ibm.com/innovation/it/ciostudy/

.[2] DRI Disaster Reovery Institute International https://www.drii.org/certification/certification.php

.[3] MySQL https://www.mysql.it/products/enterprise/replication.html

.[4] David A. Patterson, Garth A. Gibson e Randy H. Katz nell'articolo *A Case for Redundant Arrays of Inexpensive Disks (RAID)* , SIGMOD Conference (pagg. 109–116)

.[5] Google Cloud Datastore https://cloud.google.com/datastore/docs/concepts/overview

.[6] Sybil Attack http://www.math.cmu.edu/~adf/research/SybilGuard.pdf

.[7] Alessandro Panconesi, “Big Data in a small world”, Google Developers Live, https://www.youtube.com/watch?v=1smbf0eVS2I

.[8] Jaggi, S. ; Langberg, M. ; Katti, S. ; Ho, T. nell'articolo *Resilient network coding in the presence of Byzantine adversaries* , Browse Conference Publications > INFOCOM 2007. 26th IEEE Inter

.[9] Von Bayer Hans Christian *Informazione. Il Nuovo Linguaggio Della Scienza*. Dedalo Ed. La Scienza Nuova.

.[10] Elmasri-Navathe *Fundamentals of Database Systems,* 7th edition Ed. Pearson.

.[11] Matteo Golfarelli, Stefano Rizzi *Data warehouse,* Ed. McGraw-Hill.

.[12] Google Flu Trends Project https://www.google.org/flutrends/about/

.[13] Brachman & Levesqe, *Knowledge representation and reasoning,* Ed. Morgan Kaufmann.

.[14] Bruce Wong – Christos Kalantzis, *A State of Xen - Chaos Monkey & Cassandra,* Netflix Techblog, http://techblog.netflix.com/2014/10/a-state-of-xen-chaos-monkey-cassandra.html

.[15] Margaret Rouse, http://searchdatacenter.techtarget.com/definition/resiliency

.[16] Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach Mike Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber *Bigtable: A Distributed Storage System for Structured Data*. Google, Inc.

.[17] "Q316 Letter to shareholders". *Netflix*. Retrieved 17 October 2016.

.[18] Larry Page and Sergey Brin "Google Corporate Information". Google, Inc. Retrieved February 14, 2010.

.[19] *Software Piracy on the Internet : A Threat to your security*,, the Business Software Alliance, October 2009.

.[20] "*What are crackers and hackers? / Security News" . www.pctools.com*. Retrieved 2016-09-10.

.[21] Raymond, Eric (25 August 2000). "*The Early Hackers*". *A Brief History of Hackerdom*. Thyrsus Enterprises. Retrieved 6 December 2008.

.[22] Katharina Krombholz, Dieter Merkl, Edgar Weippl. "*Fake Identities in Social Media: A Case Study on the Sustainability of the Facebook Business Model*". Journal of Service Science Research (2012) 4:175-212 DOI 10.1007/s12927-012-0008-z, Published: 31 December 2012

.[23] Renee Shipley "*What is a Sybil Attack?"* Top Ten Reviews / Software / Articles

.[24] https://en.wikipedia.org/wiki/Sybil_attack

.[25] David Richards "*Big Data and the Cloud: The democratization of data in the cloud"* Network World from IDG. http://www.networkworld.com/article/3090084/cloud-computing/the-democratization-of-data-in-the-cloud.html

.[26] datascience@berkeley, Berkeley, University of California , https://datascience.berkeley.edu/about/what-is-data-science/

.[27] David Lazer,Ryan Kennedy, Gary King, Alessandro Vespignani "*The Parable of Google Flu: Traps in Big Data Analysis*", www.sciencemag.org, SCIENCE, Vol 343,14 March 2014, http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf

.[28] http://www.google.org/flutrends (2009) Google Flu Trends. Google.org. Available: www.google.org/flutrends. Accessed 2011 July 28.

.[29] CDC - Seasonal Influenza (Flu) - Flu Activity & Surveillance. Available: http://www.cdc.gov/flu/weekly/fluactivitysurv.htm. Accessed 2011 July 28.

.[30] Samantha Cook, Corrie Conrad , Ashley L. Fowlkes, Matthew H. Mohebbi, "*Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic*", Published: August 19, 2011 http://dx.doi.org/10.1371/journal.pone.0023610

.[31] David Lazer and Ryan Kennedy, "*What We Can Learn From the Epic Failure of Google Flu Trends*", https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/

.[32] Fred O'Connor (20 August 2015). "*Google Flu Trends calls out sick, indefinitely*". *PCWorld*.

.[33] "Flu prediction project by the University Osnabrück and IBM WATSON". http://www.flu-prediction.com/

.[34] J Schumacher, T Wunderle, P Fries, F Jäkel, G Pipa, "*A statistical framework to infer delay and direction of information flow from measurements of complex systems*"*Neural Computation*. **27**: 1555–1608. doi:10.1162/NECO_a_00756.

.[35] Stuart J. Russell and Peter Norvig, "*Artificial Intelligence A Modern Approach*", Prentice Hall, Englewood Cliffs, New Jersey 07632.

.[36] Russell & Norvig 2003, "*The intelligent agent paradigm*", and Russell & Norvig 2009.

.[37] R. Vilalta ,C. V. Apte, J. L. Hellerstein, S. Ma, S. M. Weiss, "*R. Vilalta C. V. Apte J. L. Hellerstein S. Ma S. M. Weiss*", IBM SYSTEMS JOURNAL, VOL 41, NO 3, 2002, https://pdfs.semanticscholar.org/c2de/34c36ed882c0bcc13a730ed09693cd660582.pdf.

.[38] Margaret Rouse, "*machine learning*", http://whatis.techtarget.com/definition/machine-learning

.[39] Rob Schapire, "*COS 511: Theoretical Machine Learning*", Lecture #1 February 4, 2008, http://www.cs.princeton.edu/courses/archive/spr08/cos511/scribe_notes/0204.pdf.