Data Driven Ecosystem – Perspectives and Problems

HANNU JAAKKOLA, Tampere University of Technology, Pori Department
JAAK HENNO, Tallinn University of Technology
JARI SOINI, Tampere University of Technology, Pori Department

Our society and business ecosystem is becoming data driven. The value of data is becoming comparable to the value of physical products and becoming an important source of business. Open data itself is seen as a meaningful source of new business, especially for small and medium-sized companies. Open data is purposely aimed at being public. In addition, there is a lot of data used as if it were public – more or less without permission. In addition, the ownership of data has become unclear – the data related to an individual is no longer under the control of the persons themselves. However, declaring data sets to be open and/or allowing access to qualified users does not yet make data useful in practice. On the contrary, this often creates opportunities for misuse and dangers regarding personal security.

Categories and Subject Descriptors: **E [Data]; H.3 [Information Storage and Retrieval]**; **H.5. [Information interfaces and Presentation]**; **K.5 [Legal Aspects of Computing]; K.6 [K.6 Management of Computing and Information Systems]; H.1.2 [Models and Principles]**; **K.6.5 [Security and Protection]** *Invasive software, Unauthorized access* - **K.7.4 [Professional Ethics]**: *Codes of ethics, Codes of good practice;* **K.8 [PERSONAL COMPUTING]**: *Games;*

General Terms: Data, Open Data, Information, Authentication, Invasive software, Unauthorized access, Codes of ethics

1. INTRODUCTION

Our societal and economic ecosystem is becoming data driven at an accelerating speed. In this context, we are not referring to the concept of the "information society" but the importance of data, or to be exact, the cultivated form of it – knowledge based on intelligent data integration – as a key element in the growth of business and welfare of societies. Data possesses business value and changes the traditions of earning models; companies like Facebook and Twitter own huge amounts of user-related profiled and structured data that is valuable in targeted marketing activities or in finding people filling the requirements of a certain kind of profile. In addition, the discussion threads of these services are providing APIs that make the data streams (of public user profiles) more or less open for data analytics; the connection networks (who is connected to whom) are also reasonably easy to analyze as well. To conclude – your value in these (social) networks is your data - not you as a person or contact.

Data is the driving force behind most (future) services and applications. The fast growth of certain innovative businesses is based on data, network infrastructure, and mobility – even in the case of physical products, they are the ultimate source of business. The beneficial use of (social) networks provides a means for communication and availability of potential collaborating (business) partners. In physical products certain properties are built-in and certified – e.g. safety, quality, suitability for use; this is not always true with data. There is also a similarity between data and physical products – e.g. both can be stolen or used in the wrong way or in an illegal / unexpected context. Databases, contact information, or personal profiles are valuable for criminals. Every day we encounter news related to cyber attacks and cyber threats, as well as problems caused by defects in information systems and hijacking of computing resources for illegal or criminal use. New innovations, like IoT, will cause new types of problems: autonomous devices interacting with each other without human control – stories about cyber attacks caused by network-connected devices are already a reality. The global character of cloud services also provides several sources of problems – some related to the safety of data repositories, some to the ownership of the data, and some in processes used to solve disagreements in legal interpretations (e.g. which law is used in globalized implementations). Insurance companies have also noticed new business

Author's address: H. Jaakkola, Tampere University of Technology, Pori Department, email: hannu.jaakkola@tut.fi; J. Henno, Tallinn University of Technology, email: jaak@cc.ttu.ee; Jari Soini, Tampere University of Technology, Pori Department, email: jari.o.soini@tut.fi.

In: Z. Budimac, M. Heričko (eds.): Proceedings of the 4th Workshop of Software Quality, Analysis, Monitoring, Improvement, and Applications (SQAMIA 2015), Maribor, Slovenia, 8.-10.6.2015. Also published online by CEUR Workshop Proceedings (CEUR-WS.org, ISSN 1613-0073)

opportunities: different data-related insurances are available that cover loss of data, losses caused by service attacks, etc; in a way, data has become concrete and a physical asset.

Our paper studies the essence of data from different points of view. Section 2 of this paper is focused on data driven changes. We will approach this topic by considering the data driven ecosystem changes driven by hyperscalability. Data-related trends are included in this discussion. Starting with the characteristics of data (Section 3) and continuing with issues related to open data. Section 4 concentrates on problems related to data driven changes. Section 5 concludes the paper.

2. DATA DRIVEN CHANGES

2.1. New ecosystems and hyperscalability

In his blog, Omar Mohaut [2015] has introduced the term "hyperscalable" to point out the importance of data in the growth of business. The only way companies that create physical goods have to scale up their business is increasing productivity (industrialization) and growth of the market. This kind of growth is capital-intensive and the limits are reached quite fast. He lists a set of new- generation companies: Spotify, Skype, Square, PayPal, Facebook, Snapchat, Instagram, Airbnb, Pintrest, Uber, Twitter, Netflix, Kickstarter, Eventbrite, Dropbox, Evernote, BlaBlaCar, Whatsapp, and Booking.com. Their business is based on scalable models and they serve millions of users with very small teams of employees. The growth of their business is not dependent on people in a traditional way, but supported by someone else's assets as a free lever. Skype has 1,600 employees to operate 40% of international telephone traffic; Airbnb has 600 employees to offer 500,000 rooms for rent without any investment (Hilton, as the biggest hotel chain in the world, has 300,000 employees to operate a hotel business of 680,000 rooms). In Whatsapp, 30 engineers support the message delivery of 7.2 trillion of messages per year (the total amount of traditional SMS messages is 7.5 trillion).

Regarding business based on physical goods, Mohaut compares traditional shopping and e-commerce. E-commerce implements a scalable business model: the website that runs 24*7 all year round and can be reached from all over the world is scalable (data-based). However, physical sales never reach such coverage, because humans as salesmen are not scalable, nor are the buildings needed for stock and shopping centers. A good example of an improved e-commerce model is Alibaba, which handles the data related to goods in the role of broker. As a scalable business concept in real physical goods, Mohaut mentions the concept of franchising: instead of delivering your goods everywhere, you rent out the business concept and brand.

To summarize – what is a hyperscalable business?:

1. A hyperscalable business model is based on intangible assets
2. A hyperscalable business model requires (information) technology as a lever
3. A hyperscalable business model uses the Internet as a free distribution channel

A business is hyperscalable when it "offers value at a near zero cost simultaneously to millions of users with a disproportionately small team." This business model is not based on the traditional factors of production in economics: land, labor, and capital but on the intelligent and beneficial use of free resources (data, Internet, social networks).

2.2. Trend Analysis as an Evidence

The studies provided by several market analysis companies provide evidence for the progress discussed above. The analysis results also point out the importance of data and (mobile) networking as the driving forces of this progress. SDTimes [2015; 2015a] has analyzed the reports of two leading companies- IDC and Gartner Group. The main trends listed confirm the growing importance of data and the Internet as key factors in progress and changes. The items gathered and combined from these trend lists cover:

1. New technology will take over the market. Growth is focused in 3rd Platform Technologies - mobile devices, cloud services, social technologies, and Big Data.
2. Wireless data growth. Wireless data will balloon to 13% of telecommunications spending. The role of mobile terminals is also changing from speech to data: according to (Finnish) statistics

(Tekniikka & Talous, March 20th, 2015),the average mobile terminal transfer data is 169 MB per day and 5 GB per month; the number of phone calls decreased by 3% in one year (2013-2014) and the number of SMSs by 16%.

3. Cloud services. PaaS, SaaS and IaaS services will remain a hotbed of activity; the highest growth (36%) is projected for IaaS adoption.
4. Big Data and analytics. In addition to the traditional structured and non-structured data, video, audio and image analytics will have growing importance. Data-as-a-Service will forge new Big Data supply chains focused on commercial and open data sets. The IDC (2012)study "Digital Universe" reported fast growth in the amount of data applicable for open analytics (Big Data),which is expected to grow from 130 EB (ExaBytes = 1018 Bytes) in 2005 to 40,000 EB in 2020.
5. The Internet of Things (IoT). The predictions identified IoT as one of the most important factors for growth of the 3rd Platform. IoT is also based on mobile communication technologies to an increasing extent. According to the current (Finnish) statistics, (Tekniikka & Talous, March 20th, 2015), 9.5% of all mobile devices are used by autonomous collaborating devices (other than mobile terminals). The IDC [2012] study "Digital Universe" predicted/ forecast?? the growth of data produced by autonomous devices: the percentage of such data is expected to grow from 11% to 20% between 2005 and 2020, indicating the breakthrough of the Internet of Things (IoT) technology.
6. Cloud services will become the new data center. Data centers are undergoing a fundamental transformation, with computing and storage capacity moving to cloud, mobile and Big Data-optimized hyperscaled data centers operated by cloud service providers.
7. Security. IDC approaches this important issue with reference to 3rd Platform-optimized security solutions for cloud, mobile and Big Data. It covers mechanisms including biometrics on mobile devices and encryption in the cloud, as well as threat intelligence emerging as an essential Data-as-a-Service category of enterprise-specific threat information. Gartner points out the importance of risk-based security and self-protection. All roads to the digital future lead through security. Because it is impossible to provide a 100% secured environment, there is a need for more sophisticated risk assessment and mitigation tools. Every app will need to be self-aware and self-protecting.
8. Ubiquitous Computing: The growth of importance of mobile devices will continue; organizations have to focus their services and applications on diverse contexts and environments.
9. Advanced, Pervasive and Invisible Analytics: There is a need to manage how best to filter the huge amounts of data coming from the IoT, social media, and wearable devices. Analytics will become deeply but invisibly embedded everywhere.
10. Context-Rich Systems: Applications are able to understand their users and are aware of their surroundings.

The trends point out the importance of mobility, cloud-based solutions and Big Data analytics. An additional factor that has an impact on future life is the easy reachability of the masses – (social) networking and its beneficial use in diverse activities. The 3rd Platform covers millions of apps, billions of users and trillions of autonomous things. Innovative accelerators are robotics, natural interfaces to services, Internet of Things, cognitive systems, and advanced security. Networks, mobile and wireless data transfer and the increasing importance of cloud services are potential sources for increasing security problems – theft of data, stealing of identity, cyber attacks, etc. The use of integrated data (as knowledge) uses diverse sources of data – open and closed. The availability, easy access, quality, and reliability of data will have increasing importance. (OR are growing in importance)

3. DATA CHARACTERISTICS – OPEN DATA AS AN OPPORTUNITY

Open Data is "data that are freely available to everyone to use and republish without restrictions from copyright, patents or other mechanisms of control" [European Union 2014]. This refers to such data sources that are open to the public by the "owner's" will. "Freely available" is implemented by providing a published interface (API) as access to the data, which is "raw" and needs further processing into useful

form. "Without restrictions" indicates fully free use of the data. In practice, there are restrictions defined by different levels of licensing.

Figure 1 (left side) illustrates the different data categories. In our classification, the term 'Small Data' is used to cover all possible data repositories – closed and open. The term 'Big Data' illustrates the part of all data that is available for analytics and which provides access for intelligent analysis tools. Openly Available Data covers such data sources that are available in networks and provide a means for monitoring its content. Part of this availability is due to a lack of or insufficient security and part is on purpose. Data sources in this category cover e.g. road cameras, weather stations, technical devices connected to the Internet, www pages, data streams, and the content of social media services. The interpretation of Open Data is explained above. An additional item in this figure is "My Data." It describes data that is in someway related to a person. By default its ownership is expected to belong to a private person, because it handles his / her private property. However, an increasing amount of this is collected by such information systems that are no longer under the control of an individual: social media services, client and membership cards, location services, etc.

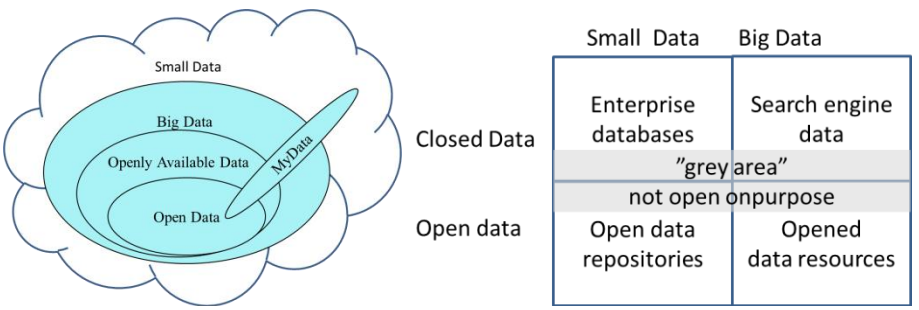

Fig. 1. Data Categories

The left side of Figure 1 approaches the classification from a different point of view. It points out the existence of data that is not open on purpose (by the data owner). This gray area data belongs mainly to the category of openly available data. It would also include such data sources that are available because of insufficient security (by accident) or used against the expectations of the data owner (e.g. www page contents).

The authors of this paper have handled the role of Open Data as a source of future business in their paper [Jaakkola et al. 2014; 2014a]. These findings are summarized below. The starting point is the Digital Agenda of the European Commission [European Union 2014]. The potential of Open Data in the EU area delivers the following benefits:

- Public data has significant potential for re-use in new products and services. Overall economic gains from opening up this resource could amount to € 40 billion a year in EU direct business volume (and € 140 billion indirectly).
- Addressing societal challenges – having more data openly available will help us discover new and innovative solutions;
- Achieving efficiency gains through sharing data inside and between public administrations;
- Fostering participation of citizens in political and social life and increasing transparency of government.

According to our studies, most of the business cases are related to marketing, better understanding of the business environment, availability of economic data related to clients and competitors, and the opportunity to develop context-sensitive services (context is based on the results of open data analysis). Weather and map data were seen as important sources. A lot of potential is also built in MOOCs (Massively Open Online Courses) in education and industrial training.

However, there are also problems. Deloitte Analytics [2014] analyzed 37,500 datasets opened in the UK (data.gov.uk; www.ons.gov.uk; data.london.gov.uk). The study shows the contradiction between the supply and demand of open data. Governmental data are usually collected for official purposes, not for pre-planned business use. The motivation to collect it comes from legal issues. This easily leads to a

situation where the interface (API) to the data becomes complex and the structure of the data is not suitable for effective and beneficiary reuse.

It is also a fact that Open Data is not open without restrictions. The most commonly applied licensing systems in open data are Creative Commons (CC) Licences (http://www.creativecommons.org). The other licensing systems, Open Data Commons (ODC) Licences (http://www.opendatacommons.org/) and Open Government Licences (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/) are identical to CC. Common to all of the above-mentioned licences is that they expect the user to acknowledge the source of the information by including an attribution statement specifying the information provider.

4. PROBLEMS

4.1. Malware

Along with the growth of the Internet (currently approx. 40% of the world population already has an Internet connection [Internet Users (2014)]), the misuse of data available on the Internet has also grown. The Independent IT Security Institute AV-TEST registers over 390,000 new malicious programs every day; in 2014, 140000000 new malware items were discovered [Malware Statistics 2015]. The annual damage to the global economy from cyber crime in 2014 is estimated to be 445 USD [Net Losses: Estimating the Global Cost of Cybercrime]. This is already a measurable part of the GDP of many developed countries, e.g. 1.6% of the GDP of Germany, 1.5% of the GDP of the Netherlands and greater than the GDP of many other countries. The actual damage may be substantially higher, since online crime costs are hard to measure - companies, banks, and governments often do not report hacking or the reports are rather ambiguous.

Malware growth is far more rapid than change in any other Internet statistics. According to the McAffe Labs report, 387 new threats appear every minute [McAffee 2015]. And this is only the visible part of the iceberg - the established anti-virus (AV) products are rather weak protection against constant and rapidly increasing threats. According to a recent report from the threat protection company Damballa [Damballa 2014], only 4% of the almost 17,000 weekly malware alerts are investigated. Inside the first hour of submission, AV products missed nearly 70% of malware, only 66% were identified after 24 hours, 72% after a full week, and it took more than six months for AV products to create signatures for 100% of the malicious files used in the study.

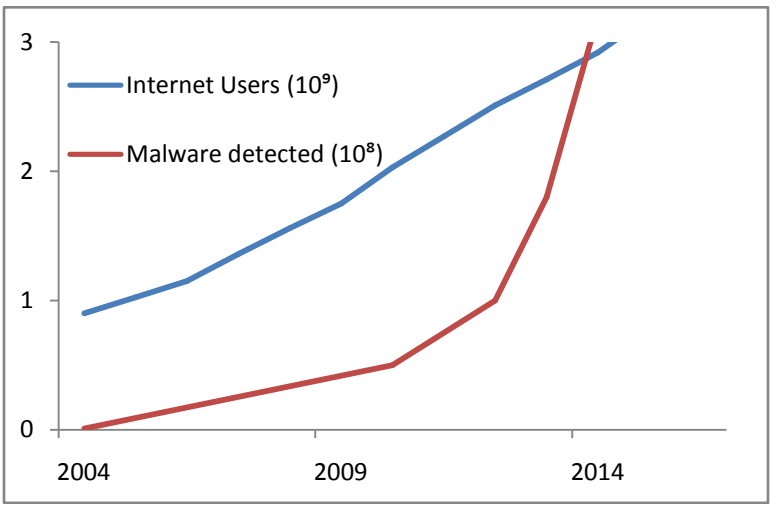

Fig 2. Growth of number of Internet users and detected malware. The exponential growth of new malware guarantees that soon there will be something for every Internet user.

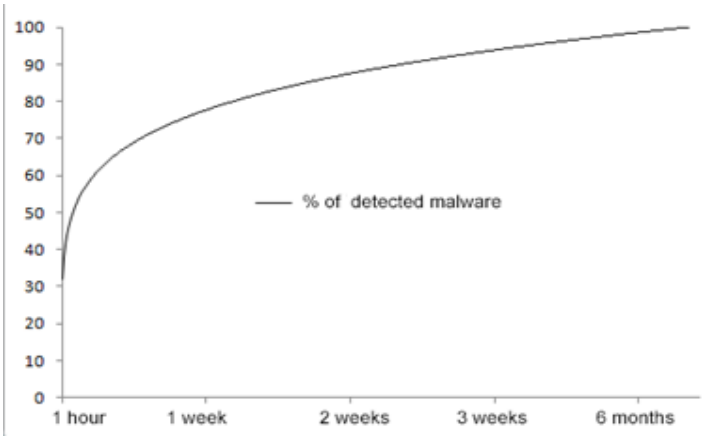

Fig. 3. Percentage of detected malware after hours of discovery. Malware completes the main part of its malicious action in the first thirty minutes after the beginning of the infection [Filiol2005], but probability of its detection in this time is less than 10%.

4.2. Personal Privacy

When you visit some Internet site to get some information, the site also gains some information about you: which web page you came from, which type of browser you are using, and your geographic location. But people often voluntarily expose themselves in some sites much more – who their friends are, what they like, where they have been recently, etc. They assume that this data remains just with this site and often even forget what personal data items they have exposed and where.

However, these items of personal data are valuable for advertisers and spammers and have created a whole new industry of collecting and marketing personal data. As a result, it has become very easy to compile presentations about privacy and security on the Internet. Just declare "there is no such thing anymore."

There are many examples. For instance, Twitter is currently planning to put trillions of tweets up for sale to data miners [The Guardian, March 2015]; nobody knows what will be revealed from this big bag of data.

Big international companies often consider that they own every piece of data they get and can handle the data just how they like, e.g. expose everything to the United States National Security Agency (NSA) surveillance program PRISM [PRISM], which collects the Internet communications of foreign nationals at several major US companies operating in other countries: Facebook, Apple, Google, Yahoo, Skype, Microsoft. When subsidiaries of these companies are registered in Europe they should also obey European laws. Under EU law, such data export to a third country is legal only if the exporting company can ensure adequate protection for such data, but NSA's PRISM program and other forms of US surveillance are the exact opposite of adequate protection. In April 2015, Facebook will be challenged by 25000 users in a Viennese Court on violation of European privacy laws [European Court of Justice 2015].

All data will sometimes become obsolete, but even obsolete data can sometimes present a threat to privacy. Facebook ended its e-mail service at the beginning of 2014 [BBC News 2014], thus currently Facebook e-mail addresses are obsolete. However, on March 20, 2015,a list containing 1642 Facebook users' e-mail addresses was released on an open Internet site. All of them are in standard form Firstname.Lastname@facebook.com or use some very similar syntax (FirstnameLastname@...), which totally exposes the real name. Unfortunately, this syntax is currently the mandatory standard in many organizations; gone are the days when you could use the address jaak@... – far more homely and revealing far less information about your real identity – where the family name is not present.

Taking at random some names from this list and making a Google query returned several websites (Instant PeopleFinders, Whitepages, Spokeoetc) where additional and often very detailed information (even without logging in and paying) were presented, e.g.

Report Includes Available Information on:***
2 matches for *****

- Current Address:…
- Friends / Family:…
- Phone Numbers:…
- Online Sellers:…
- Email Address:…
- Internet Dates:…
- Marital Status:…
- Old Classmates:…
- Location History:…
- Scammers:…
- Family Members:…
- New Roommates:…

This information was free, but for $0.85, the site promised to reveal much more.

The first line of the posting containing the list of Facebook addresses was:

"Use for Spam and if you want more msg me...."

Thus this individual knew exactly what s/he was doing - this was just a demonstration of hacking skills in order to get new customers.

Lists of e-mails are published on some sites almost daily, and it is very easy to find more such lists of e-mail addresses – just set up a crawler (using tutorials like [makeuseof 2015]) to search for some properly formatted regular expression, e.g.

[\t:="']+4[0-9]{12}(?:[0-9]{3})?

There are sites [e.g. LeakedIn] which provide such lists "just for lulz" (the plural of lol – "laugh out loud" [The Urban Dictionary 2015]).

4.3. New customs

The old generation of digital immigrants [Prensky 2001]watched cinema pictures. How very 'out'! The new brave generation of digital natives, 'internauts', instead play videogames 24/7 [WorldStar 2015] or stream their playing to others to watch on Twitch [Twitch 2015] – the modern incarnation of cinema and TV.

The old generation celebrate birthdays, marriages, births – events which have for us a deep emotional meaning. The new brave generation also have emotional events, but different. For instance, it is very important to increase the number of followers, collect 'likes' (who clicked 'like' or did something similar on your webpage) and sub/resubs (the analog of likes on Twitch). And if the number of followers/likes/resubs hits some round number, it is cause for celebration. On March 21, the following was pasted on Pastebin:

"*Hello everyone and welcome to my 100k Special stream. Hitting 100k Followers is not a small milestone and I wanted to do something crazy for it as thanks for all the support you all have shown me during my time here on Twitch/Youtube.*

… I will be streaming for 1 second for every single follower which is 100,000 seconds or 27.7 hours … every single Sub/Re-sub that happens will add 30 seconds to the total remaining time"

This is followed by the list of games (8) which will be streamed.

4.4. Who has will be given more

The Bible is a wise book – a collection of human experience from many centuries. It emphasizes some principles which are considered important for several occasions:

Matthew 13:12. Whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them.

Matthew 25:29: For whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them. …

Mark 4:25:For he that has, to him shall be given:

Luke 8:18: …. Whoever has will be given more; whoever does not have, even what they think they have will be taken from them."

etc.

While the creators/authors of the Bible had to introduce these truths from the historical experience of mankind, nowadays they can be proved.

Consider two computers (black boxes – nothing is known about their structure/functioning), which are connected.

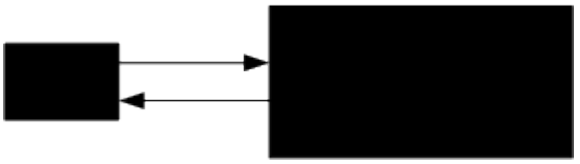

Fig. 4. Connected black boxes; one of them has far more memory.

The boxes exchange messages and store the received responses.

Since they are finite, their number of states is finite and sometimes they go into cycle, start repeating states; repeating states also occurs if some mechanism resets one/both of boxes to some initial state. Obviously, cycling happens first with the box that has less memory. Thus the box with more memory can successfully store all the responses from its little brother and when the little one starts repeating itself, it is concurred? - the big one knows exactly what the responses from the little one will be, so it can use its little brother however it likes. Instead of two black boxes, there is now only one - but with more capabilities, as the big one has the little one as a slave. To quote from Matthew, 'Whoever has will be given more, and they will have an abundance.' 'Whoever does not have, even what they have will be taken from them.' This situation has been considered by mathematicians in several papers. As long ago as 1973 the following result was proven [Trakhtenbrot & Bardzdin 1973]:

If the size of the input alphabet of an automaton (black box) $\mathfrak{W}$ is m and the size of the output alphabet - n, then for any natural number k one can effectively construct an input word $d(k)$ of length $|\mathrm{d(k)}|=]4k^2(\ln nk)m^{2k}[$ which residually distinguishes all automata with k states, i.e. any two automata with k states after getting this input either produce different outputs (they are recognized to be different) or the automata will afterwards act the same way. Since there is only a small number (length is small-power polynomial) of such words, it follows that if automaton $\mathfrak{W}$ is connected to another automaton $\mathfrak{M}_1$ with more memory and encoded to search the distinguishing word for $\mathfrak{M}$ (with number of states $c_1*|d(k)|+c_2$, where constant c_1 depends on the encoding of words of length $|d(k)|$, constant c_2 - encoding of search algorithm), automaton $\mathfrak{M}_1$ can analyze the behavior of automaton $\mathfrak{M}$, i.e. make $\mathfrak{M}$ do everything that $\mathfrak{M}_1$ wants. For the case where an upper boundary on the number of states is not known, a polynomial-time probabilistic inference procedure is described in [Rivest & Shappiro 1994].

The coming age of 'Internet of Devices' will be characterized by an increasing number of Internet connected devices with rather little memory - perfect targets which are already being exploited [Goodman 2015].In 'real' C&C (Conquer and Command) in order to perform an attack, instead of one big computer, a botnet is used - a network of Internet-connected communicating computers. Bots spread themselves from computer to computer searching for vulnerable, unprotected computers to infect and when they find an exposed computer, the machine is infected and then they report back to their master, staying invisible themselves and waiting for further commands. There are always some unprotected computers, e.g. gamers sometimes deliberately disable virus control in order to speed up gameplay. A hacker test in 2012 [Internet Census 2012] found over a million routers that were accessible worldwide. Botnets may have a few hundred or hundreds of thousands of "zombies" infected without their owners' knowledge at their disposal. Botnet creation is 'ridiculously easy' [readwrite 2013, darkreading 2014], but if you do not bother with such activity, you can rent the services of many businesses that operate almost openly or buy an executable program. And if you search, you may find somewhere a code for building a botnet or virus, e.g. code for one of the most sophisticated pieces of malware - Stuxnet - is now freely available in GitHub [Laboratory B 2014].

Fig. 5. Example product from Web store (free to download): flooder - a trojan that allows an attacker to send a massive amount of data to a specific target; user has to specify the victim's IP address, an open port, number of packets and click 'Start'. The web store offers dozens of types of malware - viruses, trojans, scanners, keyloggers, botnets, etc.

4.5. Manage your technological identity

Our identity is determined by our relations with others. Previously, relationships with other members of society were first of all physical. Gradually these physical relations have become replaced with relations based on communication technologies: mail, newspapers, radio/TV; nowadays our most important communication channel/media is the Internet. For many, increasingly their computer/laptop/iPad etc. has become an essential part of their identity – they stare at the screens hoping to get some new like/message/tweet/.. and some cannot even sleep without a mobile in their hand. And yet all the problems with this increasingly important part of our identity are acknowledged painfully. The Internet already influences the psychology of many people, who are constantly checking all their accounts, constantly uploading selfies, constantly sending SMSs. They already live in this virtual world, not in our physical world. When looking at the steady growth of Facebook, Twitter, Pinterest, etc.- the process is escalating. Google has enough memory in its servers to 'pwn' (to conquer to gain ownership [Urban Dictionary 2015a]) every single human memory. Google already understands (somewhat) natural language (queries with full sentences provide better answers than queries containing only keywords) and is rapidly improving its engine. Will we soon get all our truths from Google [ZDNet 2015], and is the whole Internet a Botnet to C&C humanity?

5. CONCLUSIONS

The paper analyzed the essence of a data driven society and ecosystem. The starting point is the phenomenon called hyperscalability. It points out the business value of data and its importance as a source of new business and activities. Open data is seen as a meaningful source of business; in addition, data that is not intended to be open is used for the same purposes. The challenges and threats related to it were discussed.

REFERENCES

BBC News (2014). Facebook quietly ends email address system. http://www.bbc.com/news/technology-26332191. Retrieved March 30th, 2015.

Chen V. H. H.&Wu Y. (2013).Group identification as a mediator of the effect of players' anonymity on cheating in online games. *Behaviour& Information Technology*, DOI: 10.1080/0144929X.2013.843721.

Damballa (2014) State of Infections Report Q4 2014. https://www.damballa.com/state-infections-report-q4-2014. Retrieved March 30th, 2015.

darkreading (2014). Researchers Create Legal Botnet Abusing Free Cloud Service Offers. http://www.darkreading.com/researchers-

create-legal-botnet-abusing-free-cloud-service-offers/d/d-id/1141418?.Retrieved March 30th, 2015.
Deloitte Analytics (2014). Open Growth – Stimulating Demand for Open Data in the UK. A Briefing Note from Deloitte Analytics. Deloitte LLP, UK. http://www.deloitte.com/assets/Dcom-UnitedKingdom/Local%20Assets/Documents/Market%20insights/Deloitte%20Analytics/uk-da-open-growth.pdf (retrieved March 31st, 2014).
European Court of Justice hears NSA/PRISM case. http://www.europe-v-facebook.org/PR_CJEU_en.pdf. Retrieved March 30th, 2015.
European Union (2014). Digital Agenda for Europe: A Europe 2020 Initiative. Open Data. http://ec.europa.eu/digital-agenda/en/open-data-0 (retrieved March 30th, 2015.
E. Filiol (2005). Strong Cryptography Armoured Computer Viruses Forbidding Code Analysis: the Bradley Virus. In Turner, Paul & Broucek, Vlasti (eds.), EICAR Best Paper Proceedings, CD-ISBN87-987271-7-6, pp.216-227.
Marc Goodman (2015). Future Crimes: A journey to the dark side of technology – and how to survive it. Random House, 464 p.
H. Jaakkola, T. Mäkinen, J. Henno and J. Mäkelä (2014). Open^n. Proceedings of the MIPRO 2014 Conference. Biljanović, P. Mipro and IEEE. ISBN 978-953-233-078-6: pp. 726-733.
H. Jaakkola, T. Mäkinen and A. Eteläaho (2014a), Open Data – Opportunities and Challenges. In Proceedings of the 15th International Conference on Computer Systems and Technologies - CompSysTech 2014 (Ruse, Bulgaria, June 27, 2014, 2014).ACM.
IDC (2012). The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf (retrieved June 12th, 2014).
Internet Census 2012. http://internetcensus2012.bitbucket.org/paper.html. Retrieved March 30th, 2015.
Internet Users (2014).http://www.internetlivestats.com/internet-users/. Retrieved March 30th, 2015.
Laboratory B (2014). Stuxnet Source Code on GitHub. http://www.laboratoryb.org/stuxnet-source-code-on-github/. Retrieved March 30th, 2015.
McAffee Labs (Feb 2015). Threats report. http://www.mcafee.com/mx/resources/misc/infographic-threats-report-q4-2014.pdf. Retrieved March 30th, 2015.
makeuseof (2015). How To Build A Basic Web Crawler To Pull Information From A Website. http://www.makeuseof.com/tag/build-basic-web-crawler-pull-information-website/
Malware Statistics (2015). http://www.av-test.org/en/statistics/. Retrieved March 30th, 2015.
Net Losses: Estimating the Global Cost of Cybercrime. Economic impact of cybercrime II. Center for Strategic and International Studies, June 2014, http://www.mcafee.com/ca/resources/reports/rp-economic-impact-cybercrime2.pdf.Retrieved March 30th, 2015.
Mark Prensky (2001). Digital Natives, Digital Immigrants. http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf. Retrieved March 30th, 2015.
PRISM (2015). http://en.wikipedia.org/wiki/PRISM_%28surveillance_program. Retrieved March 30th, 2015.
readwrite (2013). How To Build A Botnet In 15 Minutes. http://readwrite.com/2013/07/31/how-to-build-a-botnet-in-15-minutes. Retrieved March 30th, 2015.
R.L. Rivest, R.E. Schapiro (1994). Diversity-Based Inference of Finite Automata. Journal of the ACM, Vol 41 No 3, May 1994, pp. 555-58.
SDTimes, (2015), IDC's Top 10 technology predictions for 2015. http://sdtimes.com/idcs-top-10-technology-predictions-2015/. Retrieved March 26th, 2015.
SDTimes, (2015a), Gartner's Top 10 strategic technology trends for 2015. http://sdtimes.com/gartners-top-10-strategic-technology-trends-2015/. Retrieved March 26th, 2015.
Shields, Tyler (2015). "The Future of Mobile Security: Securing the Mobile Moment." Forrester Research, February 17, 2015
The Guardian (2015). Twitter puts trillions of tweets up for sale to data miners. http://www.theguardian.com/technology/2015/mar/18/twitter-puts-trillions-tweets-for-sale-data-miners. Retrieved March 30th, 2015.
Twitch (2015). http://www.twitch.tv/. Retrieved March 30th, 2015.
B.A. Trakhtenbrot, Ya.M. Bardzdin', "Finite automata.Behaviour and synthesis."North-Holland (1973).
Veracode, "Average Large Enterprise Has More than 2,000 Unsafe Mobile Apps Installed on Employee Devices." March 11, 2015.
The Urban Dictionary (2015).http://www.urbandictionary.com/define.php?term=lulz. Retrieved March 30th, 2015.
The Urban Dictionary (2015a).http://www.urbandictionary.com/define.php?term=pwnRetrieved March 30th, 2015.
WordStar (2015). Three Gamers Take Shifts Playing Video Games 24/7 For 500 Days Straight! http://www.worldstarhiphop.com/videos/video.php?v=wshhNH6R7gw53ZHz9ikA. Retrieved March 30th, 2015.
ZDNet (2015).Would you trust Google to decide what is fact and what is not?http://www.zdnet.com/article/would-you-trust-google-to-decide-what-is-fact-and-what-is-not/. Retrieved March 30th, 2015.