The Dark Side of Open Data

Matteo Mauri, Alessio Mulas, Davide Ariu

DIEE, University of Cagliari, Italy
{name.surname}@diee.unica.it

Abstract. We present a poster about a possible cyber-crime attack scenario based on data sciences, social engineering and open data. We want to raise awareness about dangers associated with the use of knowledge discovery techniques applied to open data by cyber-criminals. We hope this poster will spark interest in the topic.

Keywords: Knowledge Discovery, Social Engineering, Cyber-Crime, Open Data

1 Introduction

Humanity's streets are paved with data and as society moves forward we seem to generate an ever larger amount of them. Data is a necessary "fuel" not just for scientific progress (e.g. gene mapping, LHC,etc.) but also for communication, leisure activities (e.g. online gaming, multimedia streaming) or more important sectors (e.g. government, military, transport and enterprises). Large amounts of data would be unusable without the aid of powerful computers, as a consequence we require data to be machine readable. Western world countries are adopting a governing doctrine which holds that citizens have the right to access government's documents and proceedings to allow for an effective public oversight [1]. While transparency can help citizens controlling their governments (i.e. reducing corruption and bribery) [2] it also comes with a prize: a loss in terms of privacy. All this data is also surely useful, if not invaluable, for each kind of scam and illicit activity based on social engineering (from now onward simply SE) [3, 4]. What happens if we combine open data and SE with disciplines such as data analysis, knowledge discovery and data visualization (from now onward collectively indicated as KD)? Can we foresee a scenario where these three elements become the key of large cyber-criminal campaigns? The purpose of this paper is to demonstrate that this hypothesis is a realistic one and to provide an example of attack scenario.

2 Open Data Vs. OSINT

The rise of Internet and World Wide Web has caused an increase popularity of "open data": the idea that some data should be free to use, distribute and share by everyone without any restrictions or copyright [5]. Despite being usually associated with positive values such as "freedom", "innovation" or "opportunity", open data (from now onward simply OD) should raise some privacy concerns couple with a doubt: how much the average man or country's "decision makers" know about subjects such as data, privacy and security? From a military and government point of view there is no doubt, OD are an invaluable asset. Open Source Intelligence (OSINT) is often able to turn apparently trivial data into critical information. A famous early example is the story of how during World War II the price of oranges in Paris was successfully used as an indicator of whether railroad bridges had been bombed by Axis forces. The fact that OD and transparency are broadly perceived by people as a source of innovation and a more honest government while OSINT are associated with the world of espionage and warfare should give an hint of how much confusion or lack of knowledge there is about the subject.

3 The evolution of Cyber Crime

Open Data by themselves are already an interesting security topic but they become even more important once we take into account how cyber-crime has evolved in these last years. Long gone are the times of the lonesome nerdy looking hacker, citizen of the BBS, frantically pressing keys while burning the midnight oil over some coffee stained RFCs. These days cyber criminals are not alone anymore [6], they are organized, they are bad and they only care for one thing: profit.

Some examples: Stoyanov (Kaspersky Lab), describes Russian underground cyber-crime organizations as well structured and similar to enterprises [9]. Kharouni (Trend Micro) describes a more distributed, less structured but equally "business oriented" African cyber-crime [7]. Among the new skill-sets required by modern cyber-criminals, psychology and "data sciences" are in high demand. Psychological principles are useful to devise better scams. Cialdini's work on the psychology of influence and persuasion [4] has been already proven to be correct even if applied to computer mediated forms of communication such as emails or chat [8]. Data sciences can help criminals in several ways: finding victims (e.g. search engine misuse, Black SEO, etc.), providing support in the creation of fake web sites and social network profiles (e.g. web scraping, text analysis for chat, etc.), etc.

4 Social Engineering & Knowledge Discovery

Mitnick describes a generic SE driven attack as composed of four phases: research, developing rapport and trust, exploiting trust, utilize information [3]. It's interesting to analyse the usefulness of SE and KD applied to each one of the four phases:

- **Research** – according to Mitnick this phase should focus on open sources of information. OD are structured and machine readable making KD extremely useful for criminals. Information on a large number of victims can be collected from several sources, combined and filtered according to a desired criteria (e.g. age, gender, etc.).

- **Developing rapport and trust**. This phase is strongly based on SE. In order to gain the trust of the victim, every details is fundamental. A good message leveraging the right SE principles, sent on the best medium (e.g. chat, email, etc.) and using the best fake persona (e.g. gender, age, etc.) is bound to succeed. SE skills allow criminals to psychologically profile victims using information collected from social networks and OD. Once a victim is profiled, the criminal can easily decide what is the best strategy for the attack. SE skills can also help eliciting information from victims. KD can help the criminal solve the technical problems and automatize the process.
- **Exploit trust**. It's during this phase that the attacker asks the victim to share some knowledge or perform some action (e.g. clicking on a link, opening an attached file). KD's role in this phase is minimum, SE skills, on the other hand, can be useful.
- **Utilize information**. The obtained information can be the goal or just the starting point for a new attack. Data sciences can be useful in analysing and processing collected information.

5 An Attack Scenario

The purpose of this section is to briefly describe a possible attack scenario based on open data, knowledge discovery (and data sciences in general) and social engineering. Research – according to the law, the University of Cagliari must share as "open" all information regarding any purchase of equipment. The attacker downloads the XML file containing a date, a list of all purchased equipments and legal informations regarding the seller's company. The attacker uses seller's information as input for a common search engine and easily obtains information such as: company's web site, telephone number, legal documents. Business oriented search engines provide even maps and other intelligence. Developing rapport and trust – the attacker examines the harvested data and devises a strategy: he will send an email posing as an university employee asking for collaboration in order to solve a minor administrative problem. The seller will be told that the problem has been already solved but he is required to read the attached file and confirm that everything is correct by replying to the email. An email is considered a good attack vector since its commonly used for official communication with public institutions. The content looks realistic (i.e. is based on true facts) and is interesting (i.e. is about work). The required action looks simple and easy (i.e. just read a short file and reply). Exploit trust – the attacker uses an harvested file as attachment adding a malicious payload. Utilize information – attacker's goal is to infect victim's computer with a malware (e.g. ransom-ware, botnet campaigns, etc.).

6 Conclusions

Previously described scenario, albeit simple and "new", is not only realistic but dangerous for several reasons: it is easy to automate, can target a huge number of victims, can be applied to similar cases (e.g. procedures about contracts agreement). Following our example, represented in the poster and in these pages, if we just take into account the University of Cagliari we have the following data: University of Cagliari entrusted 6566 agreements with private firms during the year 2015, 5295 during 2014 and 2635 during 2013 with a total of 14496 in the last three years[1].

There are 93 universities in Italy and the same scenario can be applied to every public institution since the data must have the same structure (i.e. XML Schema Definition[2]).

The XML structure is known and well documented[3] .

Similar attack scenarios should be considered realistic for all UE universities and public institutions, not just Italian ones. Governments should carefully consider the impact of data sharing from a security perspective and take some actions in order to reduce the dangers associated with malicious use of this data. It should be noted that OD, SE and KW can also be successfully used to fight cyber-crime directly (e.g. www.illbuster-project.eu) or to rise awareness (e.g. www.dogana-project.eu) about dangers associated to attacks similar to the one described in this paper and not just for nefarious purposes.

Acknowledgments

This work has been partially supported by the DOGANA project. DOGANA is a project funded by the European Union Horizon 2020 framework programme, under grant agreement number 653618.

References

1. Lathrop, D.; Ruma, L., Open Government: Transparency, Collaboration and Participation in Practice. O'Reilly Media. ISBN 978-0-596-80435-0.
2. Schauer, Frederick, Transparency in Three Dimensions, University of Illinois Law Review, 2011 (4): pp. 1339–1358.
3. Mitnick, K.; The Art of Deception. Wiley, 2003.
4. Cialdini, R. B.; Influence: Science and Practice, Pearson, 2008.
5. Auer, S. R.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. . DBpedia: A Nucleus for a Web of Open Data. In Springer LNCS, 4825. pp. 722-735, 2007.
6. Broadhurst, R.; Grabosky, P.; Alazab, M; Chon, S.; Organizations and Cyber crime: An Analysis of the Nature of Groups engaged in Cyber Crime. International Journal of Cyber Criminology, vol. 8 (1), pp. 1-20, 2014.
7. Kharouni, L.; Africa a new safe-harbor for cybercriminals?. Trend Micro Incorporated Research Paper, 2013.
8. Muscanell, N. L.; Guadagno, R. E.; Murphy, S.; Weapons of Influence Misused: A Social Influence Analysis of Why People Fall Prey to Internet Scams. Social and Personality Psychology Compass, vol. 8(7), pp. 388-396, WILEY, 2014.
9. Stoyanov, R. Russian Financial Cybercrime: How it works. Kaspersky Lab Cybercrime Underground Report, 2015.

[1] http://trasparenza.unica.it/bandi-di-gara-e-contratti/

[2] http://dati.avcp.it/schema/datasetAppaltiL190.xsd

[3] http://goo.gl/WdOBBT