オンラインソーシャルネットワークにおける就職関連情報のプライバシー漏洩

于 士茜[†1,2,a)] 石川 朝久[†1,b)] フォン ヤオカイ[†1,c)]
ヴァルガス ダニロ ヴァスコンセロス[†1,d)] 櫻井 幸一[†1,2,e)]

概要：インターネットの登場以降、就職活動プロセスはインターネットを活用するようになった。ウェブで仕事や会社の情報を収集したり、履歴書やエントリーシートもパソコンで書くことが当たり前になり、利便性が著しく向上している。しかしながら、インターネットを利用した就職活動プロセスは、プライバシーの観点から意図せず個人情報が漏洩する可能性があるが、そのリスクに対する認知度はまだ低いと考えられる。特に、サードパーティが Cookie などトラッキング技術を活用して、学生や転職希望者の情報を収集している懸念が存在する。本稿では、就職関連サイトにアクセスする際、プライバシー漏洩のリスクについて定量的に評価を試み、現時点でのプライバシー保護の状況について、分析し、その動向を分析する。

キーワード：ネットワークセキュリティ，リスク分析・評価，プライバシ保護

Privacy Leakage of Job-related Information Seeking in Online Social Networks

Shiqian Yu[†1,2,a)] Tomohisa Ishikawa[†1,b)] Yaokai Feng[†1,c)]
Danilo Vasconcellos Vargas[†1,d)] Kouichi Sakurai[†1,2,e)]

Abstract: In the early days of looking for a job, individuals need to write their own resume and fax or hand to the company in person. No longer, individuals can look for jobs on websites or just fill in the electronic resume at home. However, this changed applicant process hasn't had a beneficial effect on the person who is going to find a job online. This electronic way provides a lot of privacy risks which individuals may or may not be aware of. The elements of a web page may be from another unrelated website. The information of your resume or company you are looking at is not only helping you to find a job but also give social networks an opportunity to know who you are, where you live and even where will you work in the future.
This paper investigates current privacy risks to those visiting Job-related webpages. There are 2 types of web page will be analyzed in this paper: Job-seeking website and Company homepages. It also shows the percentage of the social network at last. Our aim is to analyze how social networks collect individuals' information in job-related web page and find a way to protect individuals' privacy.

Keywords: Network Security, Risk analysis and assessment , Privacy protection

1. Introduction

With the development of Internet, it changes people's life in many ways. The way to live, the way to interact with others, and the way to find a job. And we also face a huge risk of protecting our privacy. In 2014, the Time reported, "the company will tap data it already collects from people's smartphones and other websites thy visit to improve its ad targeting. Users can opt out of such extended tracking, but they will have to visit a special ad industry website and adjust their smartphone settings to do so." [1]

In the early days of looking for a job, all the information is needed to be written in the resume and to send by fax or a letter. No longer, the job seeker can look for jobs on websites as the Internet is playing an important role in the evolution of digital technology. The study of Betsey Stevenson shows that since 2000, more and more people use the internet to find a job instead of a newspaper or personal referral [2]. It is more convenient to fill in an electronic resume at home than write a resume by handwriting. And the emerges of third-party job sites, giving job seeker more opportunities to find a job. Just fill in the electronic form, the site will find the right company for you. However, everything has two sides, it is not surprising that finding the job online has its own drawbacks. The electronic way provides a lot of privacy risks which job seekers may or may not be aware of. In the traditional way, only the company can read the privacy information of the resume. But when the information is on the webpage, it

†1 九州大学
Kyushu University
†2 九州先端科学技術研究所
Institute of Systems, Information Technologies and Nanotechnologies (ISIT)
a) shiqian.yu@inf.kyushu-u.ac.jp
b) scientia.admin@gmail.com
c) fengyk-at-ait.kyushu-u.ac.jp
d) vargas@inf.kyushu-u.ac.jp
e) sakurai@csce.kyushu-u.ac.jp

cannot be sure who is reading your privacy information online. The unrelated third-party websites or the search engines and even the company web page may collect the personal information secretly.

There are three types of sensitive information [3]: personal information, business information and classified information. Obviously, the information of the resume can be tracked back to an individual and that, if disclosed, could result in harm to that person. The job-related information includes the education information, address information and unique identifiers such as passports and social security numbers. Like the health information, job-related information has an amount of privacy information like name, gender, graduate school. And when this privacy information was collected by third-party, especially online social networks, the consequences could be disastrous. The online job-related information privacy is an issue which affects the majority Internet users who are going to find the job online. And with the rapid growth of social networks, such as Facebook, have raised a strong concern about leaking job-related information.

This paper is intended to investigate current privacy risks to those visiting job-related web pages and show the percentage of the social network at last. After analyzing how social networks collect job seekers' information in a job-related web page, there will be a tool to help job seekers protect their sensitive information.

The organization for the paper is as follows. Section 2 presents the background of finding the job online, online social network. Section 3 provides a brief overview of the related work. Section 4 describes the methodology of job-related information privacy study of online social network. Section 5 shows the results of the study. Section 6 raises a new tool to help job seekers protect their sensitive information. Section 7 is a summary and a look at future work.

2. Background

The background of finding job online and Facebook, the third-party social networking website that recently gained immense popularity is provided in this section

2.1 Finding job online

These days, online job hunting becomes a very important part of finding the job. Job seekers create the resume online. In addition, to asking for your gender, race, the questions have been specific. For example, the reference contact information or medical conditions. Generally, the job seeker cannot submit the application without providing all this data. It's like, even though you are uncomfortable supplying this information, you have no choice.

The privacy issues involved in today's new job search process are [4]:

- Collection of job seeker data

Most job sites collect the following personal information like name, address, phone number, education level, geographic location, etc. from job seekers.

Even when a job seeker simply browses through a job site looking at job ads without registering or posting a resume, the sites may still collect certain pieces of data from job seekers such as types of jobs searched for, where, and so on.

- Use of third party cookies at job and career sites

Cookies are bits of information that can be sent to a computer, and so can a company that has a banner advertisement on a page of the site. The details of third-party cookies and third-party tracking will be found in section 2.3.

Most large job search sites have relationships with one or more national advertising network, social network, including google, Facebook.com, or others.

The third-party website, especially Facebook is focused on this paper.

2.2 Social networking website: Facebook

Social networking is the practice of expanding the number of one's business and/or social contacts by making connections through individuals, often through social media sites such as Facebook, Twitter, LinkedIn, and Google [5].

As the one of the biggest online social networking website, Facebook not only has more than 1500 millions users but also has a lot of privacy concerns.

For instance, Facebook offers content providers to place a Like button on their website. The Like button is not just as a tool which allows Facebook members to indicate that they like a certain website or item on a website. It is also used to place cookies and to track and trace web users, regardless of whether they actually use the button [6]. And FaceCloak [7] is an architecture that enforces user privacy on social networking websites by shielding a user's personal information from the site and from other users that were not explicitly authorized by the user. There are several systems too. flyByNight [8] is a Facebook application designed to protect the privacy of messages exchanged between Facebook users. NOYB [9] is another system targeted at protecting user privacy on Facebook using "encryption" in a novel way

2.3 Third-party web tracking

As we analyze the third-party web tracking to decide if the

job-related websites have the risk of leaking the job seekers' privacy information. It's necessary to introduce what is third-party web tracking and how it works.

When we talk about trackers, basically, there are two types. The first-party tracker and the third-party tracker. The first-party tracker is the websites which users are looking at. They can transfer files to user's computers in order to enable websites to remember users' information, such as the username or password that have already been filled out. Third-party tracker, otherwise, can transfer files too, but not as the first-party tracker that can only focus on the trackers' website, they can gather the users' other privacy information as well.

And to learn how the third-party tracker's works, we'd like to give an example.

From Figure 1, when the job seekers are looking at the website mynavi. The first-party trackers from mynavi will remember the job seekers' information that they provide and store them in cookies. So when the job seekers login next time, they would not need to input their ID number or password, and they can see the history they looked at last time as well.

Figure 1: First-party tracking

Besides banner advertisement, there are many ways for third-party trackers to track users' privacy. For instance, the Facebook icon in the mynavi website. If the mynavi website has this icon, it means, when job seekers login this website, both my navy and Facebook store the job seekers' information in the same time, we can see the process from Figure 2.

Figure 2: Third-party tracking

And not like first-party trackers have the limit in scope to very particular data sets relevant to the website, if another website has the same third-party tracker, it means the third-party tracker can read the other website as well. The third-party trackers, unlimited in reach, are controversial and push the boundaries of privacy and ethics online.

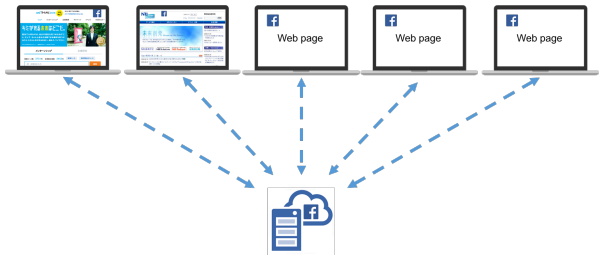

Figure 3: Third-party tracking are unlimited

From Figure 3, we can figure out that as long as this website have the same third-party tracker, they will all have the relationship to the third-party tracker. And the bigger the third-party tracker' network, the greater its potential to track users' online information.

3. Related work

Over the past few years, there has been a lot of activities on the social network and online privacy analysis. On social network analysis, some researchers have focused on graph theoretic properties of social networks [9,10,11], others have analyzed individual networks' usage patterns. And online privacy analysis often focusses on third-party web tracking. However, there has not been a detailed study of the privacy leakage about job-related information seeking in the online social network. We believe this paper is the first to measure this new workload, the job-related information, and its privacy leakage via social networks.

Facebook has been the focus of a few studies recently. In 2006, Alessandro Acquisti and Ralph Gross has shown the privacy risks of online social networks [12]. The study of Catherine Dwyer has demonstrated that Facebook members were more trusting of the site and its members, and more willing to include identifying information in their profile [13]. And Robert E. Wilson has published a study on the review of Facebook research in the social sciences, which showed the privacy risks on Facebook [14].

Prior work on third-party web tracking has three approaches to measurement: monitor network traffic, manually inspect browser state, or develop a custom tool for a specific measurement task [15]. We choose the third way: develop a custom tool for a specific measurement task to protect job seekers' privacy.

The measurement of web tracking often has two steps: first, selecting pages and second, performing automated analysis of how user data is stolen by third parties. Some study has relied

on "popular site" lists provided by Alexa company [16] [17] [18] [19], but use their own methodologies for analysis. Some study has performed comparative analyses between countries [6] and also explored general trends in tracking mechanisms [19]. And team Krishnamurthy and Wills have developed the idea of a “privacy footprint” [17]. And this team has consistently found that there are high levels of tracking on the web, including on sites dealing with sensitive personal information. However, the study of Timothy Libert on privacy implication of health information has focused on how users are tracked when they seek health information online [20].

4. Methodology

In this section, we will introduce the methodology we use to analyze job-related privacy problems. Our methodology will have two parts: page selection, social networks requests detection and analysis.

4.1 Page selection

As what we showed in section 3, there are two ways about page selection. The one is to rely on "popular site" lists and the other is to focus on users when they search on the internet.

In Timothy Libert’s study [20], the site list is made by top 50 search results for each disease. Like the health seekers, job seekers who use the search engine to find a job accounted for a large proportion. And considered to the characteristics of the job finding itself, we decide to select pages from two aspects: 1. job search sites, 2. companies' homepage. And the base search engine is Google, the language of the keyword is Japanese since the research is focused on privacy leakage of job-related information in Japan.

For the job search sites, we first compiled a list of the keyword when job seekers are finding a job online based on the data from keyword.io, the google autocompletes long-tail keyword tool. And then we collect the job search sites from the search results after input the keyword.

For the company homepage selection, we use the ranking from NIKKI, because it is the most influential website when Japanese do the analysis of the enterprise.

4.2 Social networks requests detection and analysis

To detect the third-party requests from Facebook, we used a tool named webXray [20] which is developed by Timothy Libert. This tool can identify third-party requests by comparing the domain of the web page being visited the domains of requests being made. And in our case, we only need to analyze the third-party requests that the address belongs to Facebook.

And it is also mentioned in Timothy Liber’s study that sometimes we cannot always be clear who the requested domains belong to. For instance, except the clear domain name Facebook.com, there is also an unclear domain name fbcdn.net that belong to the Facebook.

Because Timothy Libert's research is focused on the US, there are a lot of domains belong to Japan company that not being corresponded to the enterprise itself. Before we do the analysis we have updated the information about this section.

5. Findings

We scanned 222 pages which were collected by keyword of job finding and ranking from NIKKI

5.1 Third-party requests

We analyze all the 222 pages, companies’ homepage and job search site from the 222 pages. This information is illustrated in Figure 4. Of all pages, 48% initiate some form of third-party HTTP request, 46% download and execute third-party JavaScript and 18% use cookies.

It is not surprising that job search site has the most third-party requests (59%), JavaScript (55%), and cookies (29%). Although the third-party request of companies’ homepage is less than job search site, there are still nearly a half of third-party HTTP requests. Figure 4 presents these findings in greater detail.

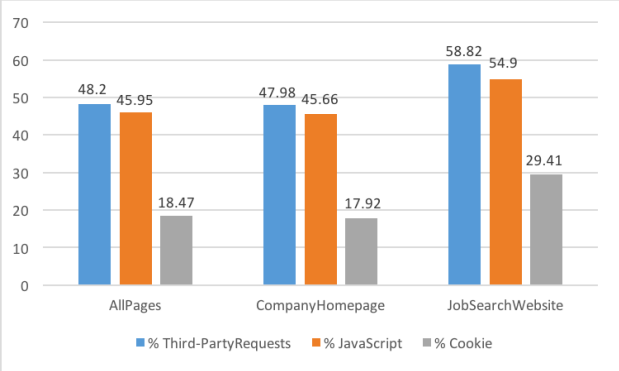

Figure 4: Proportion of the Third-Party Requests, JavaScript, Cookie

5.2 The proportion of social networks and Facebook

Given that 48% of pages make third-party HTTP requests, we still need to know the percentage of the social network and Facebook to all the third-party HTTP request.

The Figure 5,6,7 will show the details. Of all the third-party requests in all the 222 pages, almost 20% is social networks (Facebook, twitter, line etc.), although compare to the Google (40%) the social network has 20% gap, it still the

second largest proportion of the third-party requests.

It is worth noting that 40% of pages analyzed included elements which were owned by Google. Such as the traffic analytics (google-analytics.com), advertisements (doubleclick.net), hosted JavaScript (google-apis.com), videos (youtube.com).

And the second place is Facebook with 11%. It can be seen, Facebook plays an important role in social networks and has a close relationship with the job-related websites as well.

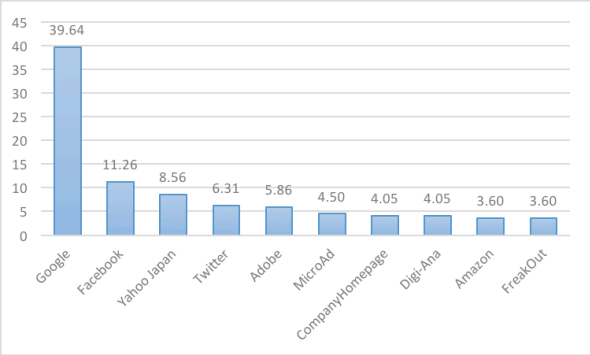

Figure 5: Proportion of Facebook in third-party requests of all pages

When we focus on the companies' homepage (Figure 6), we find that although google still has 38%, Facebook's percentage is reduced to 8%. Obviously, job seekers need to write their information or resume on companies' homepage. So the companies' homepage has a slightly higher security.

And what we can see from the Figure 6 is the third place. 7% companies collect web information from their homepage. Although they cannot be counted as third-party requests, because of the unclear domain, we need to find out if these domains are companies or not. Therefore, we collect these domains as well.

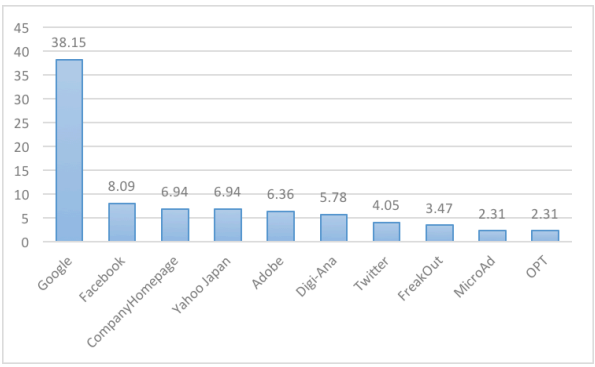

Figure 6: Proportion of Facebook in third-party requests of companies' homepage

As the Figure 7 shows that job-related websites have a high risk of protecting personal privacy. Half of the job-related websites have the third-party requests of Google, and 48% of the third-party requests are the social network. Not to mention, Facebook still has 28% in the second place.

Nowadays, more and more companies tend to submit resume on job-related websites. However, what we can see from Figure 7 illustrate that it is a very risky practice.

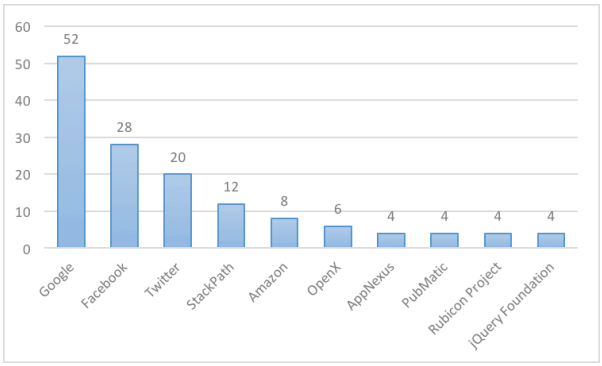

Figure 7: Proportion of Facebook in third-party requests of Job-related web page

5.3 How Facebook collects job seekers' information

Often, social media use kinds of types of file extension to make third-party HTTP request.

HTTP request with no extension, it means when the request is opened, the browser will show a blank page. And the blank page will generate HTTP requests, and the worst, it may also manipulate browser caches. And the second type is the image, for example, the Facebook 'Like' button, when a 'Like' button is showed in a job-related web page, no matter job seekers click it or not, Facebook can always record page visits. The third type is JavaScript. When requested elements are JavaScript files, these files are able to execute arbitrary code in a job seeker's browser and may be used to perform fingerprinting techniques, manipulate caches and HTML5 storage, and also can initiate additional requests.

The Table1 will provide a clear view of how Facebook collects job seekers' information.

Table 1: Types of File Extensions that Facebook use

Type	%Percentage
JavaScript	32.85
NULL	3.6
Style Sheet	1.8
Image	1.8
Dynamic page	0.9

As the Table 1 shows, what Facebook use to track job seekers occurs through JavaScript. The result is, in most of the social networks, JavaScript is the usual way to track users. JavaScript builds the social button display on the web. The 'Like' button, and also, the 'Tweet' button.

Here is how Facebook use ‘Like' button to track job seekers when they open the job-related website. First, the job seeker connects to the job-related website, and then there may be a ‘Like' button at the last of an introduction of how to write a resume. Whether the job seeker clicks it or not, the button build with JavaScript can instruct a page to open a series of URLs or download cookies, cookies have the details of what you visited, your login information, etc. They may use the cookies for targeted advertising, and the worst part is they may be sold to another company, making money off something that the job seeker never agreed to share in the first place.

6. Conclusion

As more and more people aware of the issues that online social networks may collect their privacy information when they are surfing online. There is still a field has not been noticed. When people change their way to find a job gradually, it also results in a high risk of privacy leakage of job seekers when job seekers access the job-related websites.

Therefore, we provide this paper to arose job seekers' concern about their privacy information. We show the figure about the rate of job-related websites who have third-party HTTP request. And we also pay more attention to Facebook, the famous online social network. From our analysis, as a result of Facebook occupying a large proportion of all the third-party requests.

At last, we show a table about how Facebook collect job seekers' information as well. The most usually way Facebook use is JavaScript as Facebook often show their button or icon on the job-related web pages.

7. Future Work

When job seekers finding a job online, it will involve a lot of privacy information, like personal identification, name, address, graduate school, they will associate with a unique person. And the company the job seeker is looking for, the job that the job seeker wants to find will show more details about a person.

The most important point to protect job seekers’ privacy is to avoid the correlation of the privacy and an exact person.

Several directions of future work are possible:(1) the tool can send thousands of fake information at the same time job seekers' privacy is being sent to the Facebook. (2) increase the amount of the job-related websites to improve the accuracy of the analysis.

Reference

[1] Vindu Goel. Facebook to Let Users Alter Their Ad Profiles. The New York Times. June 12, 2014.
[2] Betsey Stevenson. The Internet and Job Search. *National Bureau of Economic Research*, March 2008.
[3] Ivy Wigmore, sensitive information. http://whatis.techtarget.com/definition/sensitive-information
[4] Pam Dixon, Principal Investigator. Job Searching in the Networked Environment: Consumer Privacy Benchmarks. In *The World Privacy Forum: Job Search Privacy Study,* 2003. http://www.worldprivacyforum.org/wp-content/uploads/2012/07/wpfjobstudy.pdf
[5] Margaret Rouse. social networking. http://whatis.techtarget.com/definition/social-networking
[6] Arnold Roosendaal. Facebook tracks and traces everyone: Like this!. *Tilburg Law School Legal Studies Research Paper Series,* Available at: *http://ssrn.com/abstract=1717563*.
[7] M.M.Lucas and N.Borisov. flyByNight: Mitigating the Privacy Risks of Social Networking. in *Proc. of 7th ACM Workshop on Privacy in the Electronic Society* (WPES 2008), October 2008, pp. 1-8
[8] S.Guha, K. Tang, and P. Francis. NOYB: Privacy in Online Social networks. in *Proc. of 1st Workshop on Online Social Networks* (WOSN 2008), August 2008, pp. 49-54
[9] M. Granovetter. The strength of weak ties. *American Journal of Sociology,* 1973.
[10] S. Milgram. The small world problem. *Psychology Today*, 1967.
[11] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee.Measurement and Analysis of Online Social Networks.In *Proc. Internet Measurement Conference*, 2007.
[12] A Acquisti, R. Gross. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook, In *PET,* 2006.
[13] Catherine Dwyer, Starr Hiltz, Katia Passerini. Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and My Space. In *Americas Conference on Information Systems (AMCIS),*2007.
[14] Robert E. Wilson, Samuel D. Gosling and Lindsay T. Graham. A Review of Facebook Research in the Social Sciences. *Perspectives on Psychological Science 2012 7: 203.*
[15] Jonathan R. Mayer, John C. Mitchell. Third-Party Web Tracking: Policy and Technology. In *IEEE Symposium on Security and Privacy,* 2012.
[16] B. Krishnamurthy and C. E. Wills. Generating a privacy footprint on the internet. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement,* pages 65-70.
[17] B. Krishnamurthy and C. E. Wills. Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th international conference on World wide web,* pages 541-550. ACM, 2009.
[18] C. Castellucia, S. Grumbach, L. Olejnik. et al.Data. harvesting 2.0: from the visible to the invisible web. In *The Twelfth Workshop on the Economics of Information Security*, 2013.
[19] F. Roesner, T. Kohno, and D. Wetherall. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, pages 12-12. USENIX Association, 2012.
[20] Timothy Libert. Privacy Implications of Health Information Seeking on the Web. *In Communications of the ACM*, March 2015.