Automating Identification of Potentially Problematic Privacy Policies

Greig Paul
Department of Electronic & Electrical Engineering
University of Strathclyde
Glasgow, UK
Email: greig.paul@strath.ac.uk

James Irvine
Department of Electronic & Electrical Engineering
University of Strathclyde
Glasgow, UK
Email: j.m.irvine@strath.ac.uk

***Abstract*—Almost every website, mobile application or cloud service requires users to agree to a privacy policy, or similar terms of service, detailing how the developer or service provider will handle user data, and the purposes for which it will be used. Many past works have criticised these documents on account of their length, excessively complex wording, or the simple fact that users typically do not read or understand them, and that potentially invasive or wide-reaching terms are included in these policies. In this paper, we present our automated approach and tool to gather and analyse these policies, and highlight some interesting considerations for these documents, specifically those surrounding past legal rulings over the enforceability of some specific and widely-used contract terms — the ability for terms to be changed without directly notifying users (and presumed continued use indicates acceptance), and the protections in place in the event of a sale or acquisition of a company. We highlight the concerns these pose to user privacy and choice, and the extent to which these terms are found in policies and documents from many popular websites. We use our tool to highlight the extent to which these terms are found, and the extent of this potential problem, and explore potential solutions to the challenge of regulating user privacy via such contracts in an era where mobile devices contain significant quantities of highly sensitive personal data, which is highly desirable to service operators, as a core valuation asset of their company.**

I. Introduction

Privacy policies, terms and conditions, and other legal documents form a near-universal part of the experience of using connectivity-based services today. Virtually every website, mobile app, and even physical service provider has an agreement of this form, to which users are required to agree, in order to make use of the service. Often, however, these agreements are stated to be implicitly accepted by accessing or using a service, which gives rise to a number of considerations surrounding the validity of these agreements. Online agreements typically take the form of either a click-wrap [1], or browse-wrap [2]. These names are derived from an early form of software-related agreement, referred to shrink-wrap, whereby a user was held to have accepted a software End User License Agreement (EULA) by opening the shrink-wrap seal on the physical packaging itself [3].

The premise of such agreements was originally that it would be impractical for every user of a piece of software to individually negotiate a contract of use with the company providing the software, and record these agreements. As such, the concept of offering a standard agreement, which users accepted by using the software, was established. Such contracts are very common, and when browsing the internet, users are giving implicit consent.

Previous work has considered privacy policies, and protocols for computer-readable representations of such policies, such as P3P [4], allow web browsers to enforce a user's privacy choices. The automated processing of policies for the generation of computer-readable policies has also been carried out [5]. Despite P3P being a formally ratified standard of W3C, it has seen little traction, and has been shown to be abused by website operators [6] to work around restrictions on their sites put in place by user privacy policies. Some basic privacy-related enforcement is carried out in permissions on mobile platforms and web browsers (and these are computer-readable), although these focus simply on restricting access to data, rather than on restricting its use, which was the primary focus of the privacy policies we reviewed.

Other work has highlighted a major limitation of P3P version 1, being that policies are only accepted or rejected, without scope for partial acceptance or rejection, or feedback to the service provider [7]. Privacy is a key consideration for future network services and applications, as even the simplest of software becomes increasingly connected, as it expands onto users' mobile devices (themselves holding large quantities of personal information).

In recent years, the rise in the tendency for companies to build their businesses around the prospect of making money as a result of data gathered from the users of an (otherwise) free-to-use service has been clear. Indeed, as Bruce Schneier stated in a conference talk in 2010,"Don't make the mistake of thinking you're Facebook's customer, you're not – you're the product," [8]. With the rise in free (at point of use) services on the internet, designed to encourage users to engage with them for the purpose of gaining a larger user-base, which itself is then used as an asset [9],

This paper firstly explores some legal precedents in contract law which are relevant to the terms encountered in privacy policies. It then explores our approach to automated identification of terms of interest within policies, as well as trends identified through our analysis of privacy policies of popular services and websites. Finally, we identify potential points for future discussion and make a set of recommendations for ensuring that users' right to privacy is respected when using online and mobile services.

II. Legal Precedents

A. Alterations to Contracts

In Douglas v. US District Court ex rel. Talk America, (Case 06-75424), the Ninth Circuit Court of Appeals found that one party cannot unilaterally alter the terms of a contract. The grounds for this were that modified terms remain an offer to change the terms, until the terms are accepted. The judgement also stated "In California, a contract can be procedurally unconscionable if a service provider has overwhelming bargaining power and presents a "take-it-or-leave-it" contract to a customer — even if the customer has a meaningful choice as to service providers", referring to the 9th Circuit case of Nagrampa v. MailCoups, Inc. (69 F.3d 1257, 1283) in California. This is of interest, as a large number of technology companies are based in Calfornia, and make one of their terms of use that the agreement is covered by US law.

The Douglas ruling specifically highlighted the questions posed where a contract is modified, with the only notice given by posting a revised version. This is of interest with regard to the legal documents and privacy policies found on websites, since of the top 10 websites we investigated, all except Amazon contained a clause indicating that continued use of a service indicated acceptance of the policy changes. The original case of Douglas v. Talk America pertained to the enforceability of a change in contract to add a mandatory arbitration clause. The judgement discussed the impracticality of requiring a user to check for updates to contracts or agreements for every service they use, and highlighted that if a user was not aware of the change, the user cannot agree to the offer for change unless they are aware of it, per Samuel Williston & Richard A. Lord, A Treatise on the Law of Contracts.

When reviewing privacy policies of a number of popular websites (as discussed in Section III), we identified that Google was the only service to give a firm commitment to not reduce privacy protections without obtaining "explicit consent" of users. The majority of services made no specific commitment to directly notify users (perhaps using a direct message, or an email) of changes to policies, although many did state they may do it in some cases, although we found (for the case of Twitter, for example), that this was only if (at their own sole discretion) they decided a change was material.

Clearly there is a balance to be obtained here between notifying users of every single trivial change to a policy, and allowing changes to policies to be made without specifically contacting users. In light of the previously US-based legal rulings covering such changes, however, we believe that there is a strong argument to be made that all changes to website terms or contracts should require explicit acknowledgement and consent, to be binding.

B. Explicit Consent to Alterations

A second matter, however, covers implicit acceptance, where continued use of a service, after its terms are updated, or after a period of time, is taken by the company to indicate agreement with the updated terms. In the case of Nagrampa v. MailCoups, the court highlighted the overwhelming bargaining power of one party over the other, and that the alterations were presented on a "take-it-or-leave-it" basis. Since alterations to website policies are typically carried out as such (without provision made for users unwilling to agree to the new terms), this ruling may prove relevant when considering internet-based services, as the provider holds a position of overwhelming bargaining power (by being able to prevent a user from accessing their own data until they accept new terms). We believe that this position may well form the grounds for modifications to be deemed unfair, on account of the imbalance of bargaining power between user and service provider.

This is also inkeeping with UK unfair contracts law guidance [10], which is itself derived from a European directive, meaning that similar terms should exist throughout the European Union. We believe that contracts between users and service providers are inherently designed with an imbalance, given the service provider is able to dictate that a user accessing their own data (held on the service) must accept new and modified terms, with the user having no clear option to access their data without accepting those terms. Indeed, in some cases, users may not need to even use the service, as we found acceptance was in some cases presumed after a specific period of time. This raises concerns for services where users are not given a clear ability to remove all their data from a service (and terminate their dealings with the company). Another consideration here is whether it is fair or reasonable for a user to no longer have access to a service, simply on account of their objection to a unilateral modification to the contract, which was not in place when they accepted the agreement. Had they been aware of the intention of the provider to make this change, they may never have used it, or may have used an alternative.

C. Transfer of Data Following Acquisition

A common term seen within privacy policies and terms of service covers the ownership (and transfer of) user or customer data, in the event of the sale, acquisition or bankruptcy of the company operating the service. These terms find their origins as a result of a number of cases in the USA, where the Federal Trade Commission (FTC) successfully argued that a company undergoing bankruptcy was bound by its original promises in a privacy policy, such as to not sell private data, or to protect it [11]. Specific examples include that of RadioShack and Toysmart. While there has been considerable legal attention given to the subject, online services are recognised as being largely self-regulated, with regard to the handling of personal data (through privacy policies), which is why the FTC seeks to hold companies to account and ensure they honour their own privacy policies [12].

As a result of this, it is common to see website and service privacy policies clearly state that data may be transferred or sold as part of a sale, bankruptcy or acquisition. By stating this may happen, the company is acting within its own privacy policy, and can likely continue with the sale unhindered. Indeed, previous legal challenges brought forth by the FTC have focused on holding companies to their own privacy policies when attempts at selling user data are made, and in the absence of legally binding protections for private data, it appears acceptable for personal data to be sold as part of a bankruptcy or acquisition, provided such a provision was made in the privacy policy in question.

TABLE I. QUERY TERMS

Hyperlink Query String	Page Content String
terms	transfer
privacy	acqui
legal	merge
	amend
	modif
	notif

In order to investigate the prevalence of these kinds of terms, we present our analysis of a number of privacy policies, accelerated through the use of our toolkit, used to identify potentially interesting contact and policy terms for further analysis. These policies belong to the most popular web and mobile-oriented services.

III. AUTOMATIC PROCESSING

To gain a clear understanding of the extent to which some of the concerns highlighted in Section II are found in policies available today, we created a toolkit for the purpose of analysing these policies. The toolkit accepts a CSV list of top website domain names (the Alexa listing being one widely known example of such a dataset). In order to simplify analysis, at this point only sites using the '.com' and '.co.uk' top-level domains are considered, since these sites typically feature their legal policies in English. In future, detecting website language based on the page content would allow for this to be expanded, and would allow for non-English website policies to be processed, provided suitable policy definitions were created for the language in question.

The identified websites were then accessed on their default "www" subdomain prefix by our automated tool, and the homepage was loaded into an HTML parser, which identified all hyperlinks within the page containing a phrase relevant to legal policies. These links were then followed, and the policy documents gathered and scanned for strings likely to be of interest in investigating the transfer of personal data, or the alteration of policies. At this time, we used the phrases shown in Table I, which were themselves selected from the top 20 websites.

IV. CHALLENGES IN AUTOMATED PROCESSING

In the process of carrying out this automated processing, we identified a number of challenges when attempting to automate the process of extracting key privacy policy information. Firstly, as discussed previously in Section III, we did not at this point consider websites written in languages other than English, although the process could be applied to websites of any language, provided suitable experience was available for identifying suitable query strings. Secondly, the challenge of ambiguity in links was also identified, where a website may contain more than one link mentioning "privacy" or "terms". We identified two potential approaches to handle this scenario. One approach was to use the link with the shortest visible text, since articles mentioning the query strings or other, more generic pages about privacy tended to be qualified with other words, such as "learn more about privacy". The other was to use the link which was found last on the page, since all websites we investigated listed their privacy policy at the footer of their page, which is conventionally (but not always) listed towards the end of the HTML in a page. One other potential challenge we identified (but did not encounter on any of the most popular websites we investigated) was the use of javascript-based hyperlinks to show a privacy policy. Since these would require javascript code to execute in order for the policy to display, retrieval of the destination of the hyperlink would not show the policy. While we did not encounter this on any websites we attempted to process, it would potentially lead to the tool not detecting any relevant policies on the page. For this reason, manual supervision of the acquisition process remains necessary at this point, to ensure the policy is identified correctly.

After identifying the relevant hyperlinks within a website, we identified a number of challenges when attempting to automatically process the policies themselves. While our process of attempting to extract context of a query term by retrieving the full HTML component within which the term appeared was relatively successful, it was not always perfect. Specifically, if a sentence was split (as is common when breaking up lengthy legal statements using bullet points, or sub-sections), the subsequent information was lost. To ensure that findings were accurate for this work, we manually reviewed the relevant terms of a policy (in its regular context of the full policy). By also taking into account the variations in use of terminology, fully automated processing of these privacy policies and other legal documents remains a significant hurdle. Some relatively common words (such as merge, or modify), are by definition necessary to highlight for the purpose of identifying policy terms of interest, and are also likely to be found in other contexts (such as users being able to modify existing content), or indeed within in-page javascript content to merge multiple objects together when rendering the web page. The latter of these were able to be removed using length filtering (to remove infeasibly long lines of content), although parsing the string as javascript to verify syntax also allows these false positives to be removed.

V. KEY OBSERVATIONS

As a result of carrying out this investigation, we identified a number of key findings, which should be considered in line with the discussion regarding enforcement of terms in Section II. Specifically, our findings lead us to believe that;

- Almost all online services deem continued use of a service as acceptance of new terms or policies, irrespective of whether the user is made aware of the changes, which is potentially not in-keeping with legal precedent.
- Very few services promise to directly contact and inform individual users of changes to policies. Most say they may do this, but some services may only post a message on their service website, or perhaps only even update the policy.
- Some services say they will notify users of major changes to their policy, although this is as determined by the company itself, and in some cases the terms state that only the company may determine if a change is major.
- Service operators may in a position such that rights in a policy or contract could be found significantly

imbalanced — a user must agree to arbitrary terms to continue to access their own data on a service, and their access to that data can be terminated if they refuse to consent to the new terms.

- Many services state in their terms that user data may be transferred to another company in the event of an acquisition, merger or sale. The privacy protections experienced here may well differ from what users expect, or previously were in force.

In the case of policy updates, we also believe that, on account of the wide use of smartphone applications, not all users will encounter a notice of a change placed on the homepage of a service. For example, many users of Google services on Android will have had an account created during the Android setup process, potentially within a retail store, and may not ever use (or know their password for) the web interface for these services. Additionally, especially on mobile devices, users are often prompted to accept policies without necessarily having even seen the policy - due to screen space limitations and the lengths of these policies, the policy was often made available by way of a hyperlink, which upon clicking would open a new screen containing the policy in question.

A. Jurasdiction of Contracts

Another major consideration when reading privacy policies and other legal documents, specifically pertaining to the usage of internet services, is the jurasdiction under which contract is governed. While most policies we found stated this, we note that there can be considerable variation in legal protections for users between jurasdictions. This is particularly clear when US and European Union data protection laws are considered. While the *Safe Harbor* process allows for companies to self-declare their compliance with European law. This aims to remove the challenges faced with the EU's comparatively heavily regulated handling of private data (which is established as a "fundamental right" [13]), in comparison with the relatively hands-off approach taken in the US, where privacy protections are typically implemented voluntarily by companies, in order to prevent potential lawsuits [13].

VI. Recommendations for Privacy

As a result of these observations, we present some recommendations to improve the fairness of privacy policies for end-users. Firstly, we believe it essential that all users are made aware of all changes to a privacy policy or other legal document governing the use of provision of a service. This notification should be made directly to the user, in plain language, and in an honest and understandable way. We recognise the trade-off between annoying users, and ensuring they are informed of changes, but believe that given the legal precedents discussed in Section II, it is necessary to ensure users are definitely made aware of changes. In the meantime, we believe there would be opportunity for a free service to poll important legal documents of web services regularly, and send a regular digest, notifying users directly of changes to policies on services they indicate they use. We believe that our approach to identifying these policies automatically from a web page could be used to underpin such a service, by comparing the text regularly, and alerting users if key words were to be added or removed.

As an extension of the above, we recommend that services should also seek explicit consent when updated terms or policies are presented. This ensures that users have directly consented to the variations in terms, and that there is no presumption of consent. If a user no longer uses a service, for example, their consent may be presumed when the user was unaware of the alteration (and unwilling to agree to it). By requiring explicit consent to all changes (perhaps at subsequent logins, as some websites implement), service operators know which users have consented directly to the changes, and can separate user data which is not able to be used under the updated policies..

Similarly, we recommend that the tie between agreeing to new terms, and being able to continue to use a service which has already been available presents a question and challenge for further discussion. If a user is prevented from accessing their data until they agree to a new set of policies, they are arguably being unfairly pressured into agreeing to a change in policies or contract. This could well be the equivalent of being forced to sign an agreement under duress, which would ordinarily invalidate it. Nonetheless, we recommend this topic for further discussion, as it highlights the differing needs of service users and providers. While service providers typically wish to protect themselves from legal challenge, and ensure their users understand the ways data may be used by the provider, the user of a service wishes to understand how their personal information is used, and to exercise control over this to prevent use which they do not consent to. By being able to balance these needs (perhaps through providing access to the previously-offered service for compatibility), we believe a satisfactory outcome can be achieved, such that service providers may change policies, but users may continue to use services which they are reliant upon, without being coerced or forced into agreeing to something they are unhappy with. When personal data is being handled, the ability to challenge a change after the fact may not necessarily offer a satisfactory remedy, since the personal data may already have been shared or sold or otherwise used without the user's consent, and it cannot easily be "taken back" after the fact.

Finally, we recommend that existing standards (such as P3P) for the machine-readable representation of privacy policies should be revisited, as they offer significant benefit for those wishing to compare privacy policies between websites. For this to be effective, however, it would be necessary that service operators and websites be bound by their P3P-stated policy, such that users may allow their browser to make decisions based on website policies. If service operators believed that their human-readable policies were the canonical definition of their policies, this would potentially allow for potential abuse, where sites would place incorrect P3P policies on their site to mislead users into disclosing information under false pretences.

VII. Conclusion

This paper has discussed some of the legal and user-facing considerations of privacy policies online, as well as our approach to automatically locating them on websites, and

attempting to highlight important portions of these policies for further review. Our approach has been used to automatically flag sections of privacy policies for further review, which appear to present clauses which act against the legal precedents we explored. We presented a set of observations across our analysis of many of these policies from some of the most popular websites on the internet, and have highlighted and made recommendations with regard to specific concerns surrounding presumed or implied consent to updated policies, and of the notification of users to changes in such policies. We have also highlighted and made recommendations regarding the all-or-nothing approach to use of services, and the potential risks of data lock-in being used to force users to agree to privacy policy or other legal changes under what may be perceived as duress.

ACKNOWLEDGMENT

This work was funded by EPSRC Doctoral Training Grant EP/K503174/1.

REFERENCES

[1] C. L. Kunz, M. F. Del Duca, H. Thayer, and J. Debrow, "Click-through agreements: Strategies for avoiding disputes on validity of assent," *The Business Lawyer*, pp. 401–429, 2001.

[2] C. L. Kunz, J. E. Ottaviani, E. D. Ziff, J. M. Moringiello, K. M. Porter, and J. C. Debrow, "Browse-wrap agreements: Validity of implied assent in electronic form agreements," *The Business Lawyer*, pp. 279–312, 2003.

[3] R. H. Stern, "Shrink-wrap licenses of mass marketed software: Enforceable contracts or whistling in the dark," *Rutgers Computer & Tech. LJ*, vol. 11, p. 51, 1985.

[4] L. Cranor, M. Langheinrich, and M. Marchiori, "A P3P preference exchange language 1.0 (APPEL1.0)," *W3C working draft*, vol. 15, 2002.

[5] C. A. Brodie, C.-M. Karat, and J. Karat, "An empirical study of natural language parsing of privacy policy rules using the SPARCLE policy workbench," in *Proceedings of the second symposium on Usable privacy and security*. ACM, 2006, pp. 8–19.

[6] P. G. Leon, L. F. Cranor, A. M. McDonald, and R. McGuire, "Token attempt: the misrepresentation of website privacy policies through the misuse of P3P compact policy tokens," in *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society*. ACM, 2010, pp. 93–104.

[7] R. Grimm and A. Rossnagel, "Can P3P help to protect privacy worldwide?" in *Proceedings of the 2000 ACM Workshops on Multimedia*, ser. MULTIMEDIA '00. New York, NY, USA: ACM, 2000, pp. 157–160. [Online]. Available: http://doi.acm.org/10.1145/357744.357917

[8] D. Shane. Facebook is "deliberately killing privacy", says Schneier. [Online]. Available: http://www.information-age.com/technology/security/1290603/facebook-is-%22deliberately-killing-privacy%22-says-schneier

[9] T. Louis. (2013, 31 August) How much is a user worth? [Online]. Available: http://www.forbes.com/sites/tristanlouis/2013/08/31/how-much-is-a-user-worth/

[10] Competition and M. Authority. (2014, 5 June) Unfair terms in consumer contracts regulation explained. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/317667/UTCCR-explained-an-overview-in-an-easy-to-digest-format.PDF

[11] M. P. Heuga, G. Manglik, and C. Meyer. (2015, 13 August) Transferring customer data in an asset sale. [Online]. Available: http://www.jdsupra.com/legalnews/transferring-customer-data-in-an-asset-47838/

[12] W. E. Agin, "The new regime for treatment of customer data in bankruptcy cases," *J. BANKR. LAW & PRAC.*, vol. 10, p. 365, 2001.

[13] W. J. Long and M. P. Quek, "Personal data privacy protection in an age of globalization: the US-EU safe harbor compromise," *Journal of European Public Policy*, vol. 9, no. 3, pp. 325–344, 2002.