Interactive Info graphs as an alternative to text-based privacy statements

(Dilip Sahoo and Aaruni Upadhyay)

Introduction

Textual privacy policies that are used by most online services today are lengthy and the users often find these privacy policies uninspiring. Researchers at Carnegie Mellon University found that it will take users 201 hours a year just to read privacy policies [1]. Users are also discouraged by the length of such privacy statements, so they simply resort to agreeing to the conditions without understanding what they are consenting to [2]. In this project, we present an alternate way to present privacy policies to the reader that the she will find not only more interesting, but it will also lead to a better comprehension of the contents of the policy.

Objective

In this project, we provide a visual way to present a summary of typical privacy policy by visualizing the data flow through a directed graph. The directed graph captures the essence of the privacy policy and lets the user interact with it through nodes that can be expanded and collapsed. When a user consumes a privacy policy in this manner, it encourages her to understand various elements of the privacy policy, instead of just scrolling down and clicking on the "Agree" button.

Tools

- C# .Net, Node.JS, NW.js, Javascript, XML
- Visual Studio 2019 for writing XML Generator
- Stix-Viz (https://github.com/STIXProject/stix-viz) for XML Visualization
- Notepad++ for XML Editing
- VMWare Workstation for creating a Win 7 Development Environment

Methodology

1. We took Honda privacy policy from https://www.honda.ca/privacy as an example to base our project on. We then divided this text-based policy into 9 sections [3], namely: Collection, Sharing, Choice, Access, Data retention, Data security, Policy change, Do not track and Purpose. These sections were then broken down into individual subcategories to identify the different elements of the privacy policy and the relationship among them as shown in Table 1.

Collection	What data do we collect?	Information you provide	Name
			Address
			eMail
			Phone
			Driver License #
			Insurance
			Type and model of Honda Products and Services
			Purchase, service and/or repair history
		Technical Website Information	IP Address
			Unique Device Identifiers
			Browser Type
			Internet Service Provider
			Hardware Model
			Operating System
			Application Version
			Location information
			Referring and exit webpages
			Time and Date stamp
			Search terms
			Clickstream data
		Social Media and 3rd Party Info	Profile data
			Avatar
			Name and/or User Name
			Network connections
			Other information
		Career Information	Name
			Educational Background
			Employment Background
			Contact Information
			Job qualifications
			Jobs Interested In
			Resume
			Employment References
			Salary requirements
		Communications and Interactions	Comments
			Complaints
			Feedback
			Inquiries
	How do we collect data?	Within Honda	Honda Dealers
			Onboard vehicle systems
			Point of purchase or lease of your Honda Products and Services
			Service transactions with authorized Honda dealers
			Your customer service inquiries
			Online Honda account
			Web-tracking tools
		Outside Honda	Affiliates
			Service providers
			Third parties
			Customer service surveys
			Contest entry information
Sharing	Why do we share data?	Business Activity	Processing credit application
			Debt collection
			Purchase, finance or lease transaction
			Maintaining warranty and customer service records
			Recall campaigns
			Fraud detection and prevention
			Customer service campaigns
			Providing marketing information
		Other	Maintaining records for legal purposes
			Conducting market research
			Other legal purposes
	With whom do we share data?	Within Honda	Subsidiaries
			Dealers
		Outside Honda	Service providers
			Marketing providers
			Other third parties
			Government authorities outside Canada
			Affiliates and Partners

Table 1

Showing 2 privacy sections (Collection and Sharing) out of 9 that can be extracted from Honda.ca privacy policy.

2. We built a new application in C#. Net called "XML Generator" that takes sub-categories from step 1 above and allows the privacy policy writer to create an XML file from it. This XML file will contain the sections and subcategories and the relationship that exists among them. Figure 1 below shows the application in action.
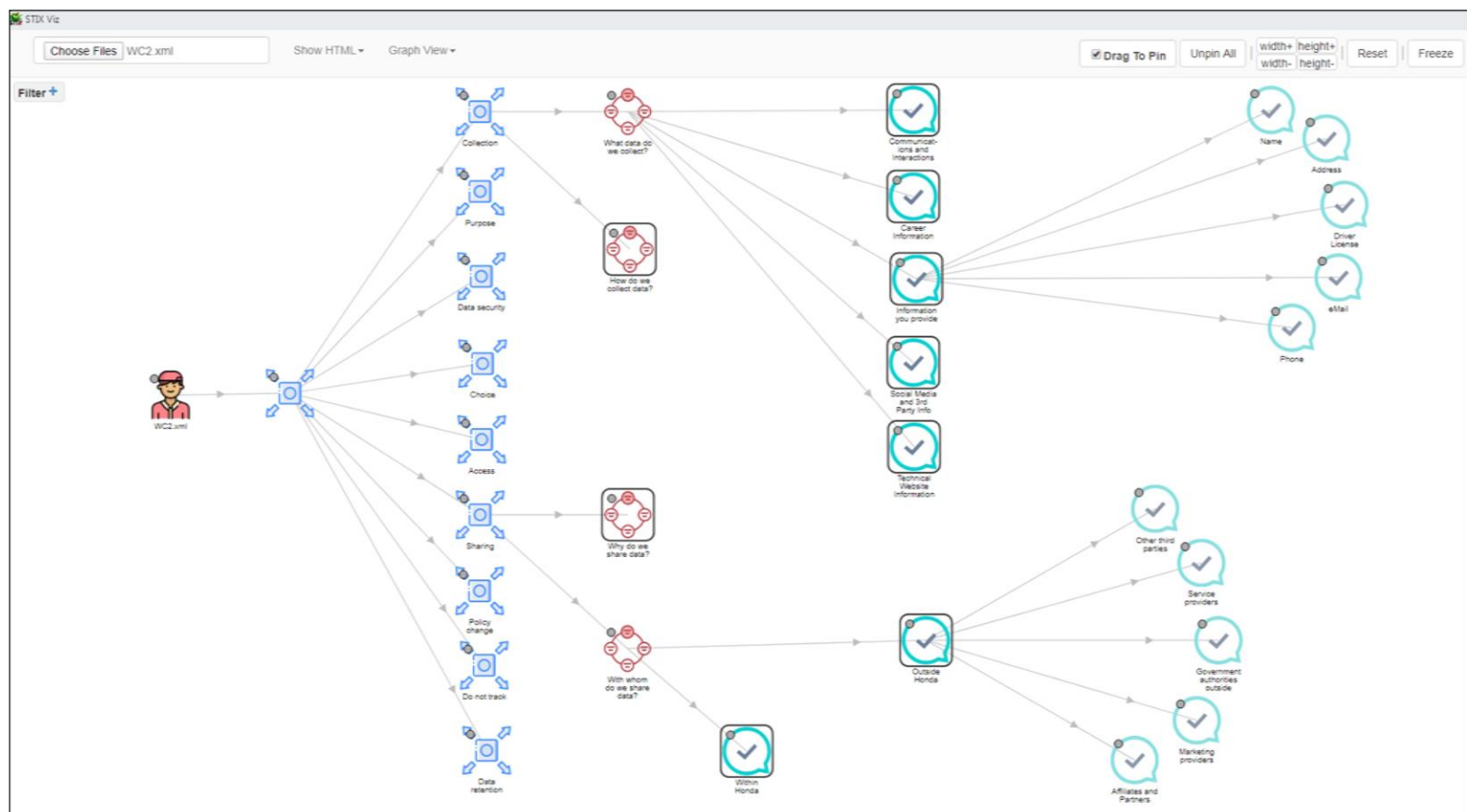
Figure 1: XML Generator Application

Figure 2: Graph representation of Privacy Policy

3. We modified the Stix-Viz application that is used to visualize threat information as a clickable node-link tree structure from XML based STIX documents. We then supplied the XML we created in step 2 above to Stix-Viz to display the privacy policy in a visual form of nodes and links as shown in Figure 2, 3 and 4.

Figure 3: Zoomed-In Top-Level Nodes

Figure 4: Zoomed-In Leaf Nodes

Results

1. The visual representation of the privacy policy was obtained successfully.
2. The XML Generator application, XML file structure and Stix-Viz application seem to be working as intended with each other.

Advantages

In our opinion using a visual representation over text-based privacy policy offers these advantages:

1. A visual representation is more appealing to interact with instead of reading a wall of text. This increases the likelihood that the user will try to understand the policy.
2. It highlights the hierarchy and the relation between different entities in the privacy policy in a much clearer and easier to understand fashion which becomes difficult in long text-based policies.
3. If desired, this visual presentation can track up to what extent has the user interacted with the policy which is not possible with a text-based approach. For example, a website owner can restrict user registration until all the leaf nodes have been clicked.
4. The drill down nature of this visualization enables a user to quickly find specific sections of the privacy policy.
5. The visual representation also allows for the gamification of a privacy policy. For example, points could be awarded for each click on a node and a coupon code could be revealed once the user has clicked on more than 90% of the nodes.
6. Our solution is scalable i.e. it can be used for both short and large privacy policies by generating XML as necessary.
7. The use of XML makes it easier for using different applications and technology. For example, we can use some other visualizer as long as it can accept an XML file as an input.

Disadvantages

1. It requires additional effort on the creator's part to translate the privacy policy into simpler elements that can then be used to create XML.
2. Very large policy policies could make the visual representation seem cluttered due to the presence of a large number of nodes. This could be resolved by careful selection of the level of detail to be displayed and by breaking the policy across multiple sub-graphs.
3. It may be hard to capture all the semantics of a privacy policy in this visual form especially when it is crucial from a legal point of view.

Future Work

We have the following suggestions for researchers to build upon our work:

1. Add feature to automatically extract the nodes from an existing text-based privacy policy instead of using the "XML Generator" application. This will avoid a lot of rework by the policy creator.
2. Use gamification to make the policy more exciting for the end-users.
3. Add the functionality to provide more details when the user clicks on the leaf nodes. These details could be in the form of text or additional links for the users to navigate to.
4. Rewrite the XML visualizer to remove the dependencies from the format underlying the Stix-viz application. This will make the XML smaller in size by removing the unnecessary elements.

References

[1] I/S: A Journal of Law and Policy for the Information Society, vol. 4, no. 3 (2008), 543-568

[2] Jonathan A. Obar & Anne Oeldorf-Hirsch (2020) The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services, Information, Communication & Society, 23:1, 128-147, DOI: 10.1080/1369118X.2018.1486870

[3] Jasmin Kaur, Rozita A. Dara, Charlie Obimbo, Fei Song & Karen Menard (2018) A comprehensive keyword analysis of online privacy policies, Information Security Journal: A Global Perspective, 27:5-6, 260-275, DOI: 10.1080/19393555.2019.1606368

Attribution

1. SVG Icons for use in Stix-Viz application were obtained from www.FlatIcon.com
2. Stiz-viz uses NW.js from www.nwjs.io to run Node.JS libraries