Consolidated Review of

Profiling High-School Students with Facebook: How Online Privacy Laws Can Actually Increase Minors' Risk

1. Strengths:

The key idea of this paper is simple but effective, beginning with core minors. It can reveal most of minors protected by the COPPA law. The strongest point of this paper is using the real data obtained from three high schools. It quantitatively verifies authors' methodologies beyond the simple simulation.

I appreciated how careful the authors handled the ethical issues related to the study and information they collected.

Moreover, what-if scenario assuming the situation without the COPPA law is the great demonstration. This Section 6 gives another view for understanding the impact of core minors who lied their ages. I also like Section 7, countermeasures. Since COPPA-less world is infeasible, but the partial solution to protect minors' privacy leakage is valuable.

2. Weaknesses

I completely understood the value of datasets authors studied and the difficulties to obtain. However, from the perspective of the statistical significance, I am not sure how we can generalize the result based on only three high schools. If authors add a section to discuss the generality of this work, it would be great.

I hope the authors better explain the implications of their paper to the IMC audience, which may has less expertise in the policy area. For example, I don't have the expertise to vet whether the findings are a real threat to public safety or whether the conclusion about COPPA is a valid one. Perhaps references to work in public policy or other conferences could help justify the threat better.

As a layman, I don't completely agree with the conclusion that COPPA "caused" this problem. From a legal perspective, couldn't one also blame Facebook for failure to validate members' ages? This is another area that the authors could present some citations to improve the legal/policy clarity for us non-experts.

Although this work is definitely great, I'm not sure whether this paper fits IMC.

I wish the paper could evaluate information leakage in a more depth analysis. The results in this paper are mainly about how many students they could obtain information. It would be interesting to see how much information they can obtain for each student (beyond what shown in Table 5 and Table 6)

The paper is extremely verbose, and diagrams can easily improve the presentation of the methods.

3. Comments

This is an intriguing study into an important problem. The findings that so many minors lie about their age in order to get access to Facebook and thus are vulnerable to this attack is a bit shocking even if it just confirms conventional wisdom.

That said, I'm not sure what the Internet measurement community can do with these results or the methodology. The findings seem more appropriate for a policy conference. Do you think the methodology would be useful to study other vulnerabilities or aspects of social networks? That would make it more relevant if technical practitioners could use it some how.

While the paper implicates this problem to unintended consequences of COPPA, I would lay more of the blame on Facebook for not properly validating members' ages. While this perhaps would not be in Facebook's financial interests, it seems to be that these results demonstrate that by not doing this, they are in fact not meeting the spirit of the COPPA law. To make an analogy, suppose a law gave an incentive for people to lie to banks (many such laws exist). If banks turned a blind eye to such lying since it was in their financial interest, I doubt one would conclude that they are blameless.

The paper is also very verbose. I suggest to the authors that they should sharpen the text and reduce the paper length. It takes roughly 4 pages just to reach the method. I would propose to reorganize the paper in a way that brings substance earlier. There is a single idea in the paper - leveraging kids lying about their age - and the rest is just circulating around this idea. The idea is nice, but it is not a measurement paper, nor does it give us some generic method that can be applied elsewhere. Perhaps that is why it is so verbose. And I agree that it is likely to be out of scope.

A recent related work that you did not cite: "Private traits and attributes are predictable from digital records of human behavior" in PNAS 2013, where the authors used several features from a Facebook profile to infer various hidden aspects of the users, one of which is the age. However, their work used a Facebook app that connects to the person, and used her private data to infer various non-disclosed properties.

As the extension of Section 7, I think additional feasible solutions make this work more complete. For example, as a very simple solution, Facebook can automatically exclude some users from search results if they have many minors as friends. Does it work? I think authors have enough data to emulate/simulate this sort of new protection schemes.

4. Summary from PC Discussion

There were several round of discussions of this paper during the TPC meeting. The main concerns are: 1. Whether the paper is considered as a measurement paper and within the scope of IMC and 2. If and how IMC community can benefit from the methodology or results of this paper. At the end of the discussion, most people feel that it will be an interesting paper to be included in the IMC program as it identified an unexpected information leaking problem with kids lying about their ages on Facebook.

5. Authors' Response

We would like to thank the reviewers for their thoughtful comments. This paper collects information from an OSN and shows how this information can be used to infer sensitive information about minors. It shows that even a very cautious

minor can be profiled through inference. Before submitting the paper to IMC 2013, we asked the PC chairs if a privacy study involving Facebook data was within the scope of IMC 2013, and the answer was yes. One of the reviewers felt the paper was verbose; but the same reviewer asked for more motivational discussion! The topic is complex with many nuances, so a fair amount of explanation is needed. We nevertheless trimmed lightly a few paragraphs. We agree that the problem could be greatly alleviated if Facebook validated the ages of its members. But the COPPA law does not explicitly require this (or else, online services would be doing it). We feel that information leakage has been thoroughly quantified in this paper. To highlight this, we moved the section on "extending the profiles" from the appendix into the main body of the paper. Actually there are two main ingredients to the attack: leveraging the minors who lie about their age; and reverse friend look up. We corrected all the minor typos and improved the figures, as the reviewers suggested. We also included the suggested reference.