[6] Modelling the Cybercrime Cascade Effect in Data Crime

Maria Grazia Porcedda (Trinity Colege Dublin) and David S Wall (University of Leeds).

Reviews Phase 1

Review 1

Time: Jun 02, 20:27

Overall evaluation: 1 (weak accept)

This submission requires significant revision before publication. But there are enough positive elements in it to suggest that it should be persevered with, if both the author(s) and PC feel there is sufficient time to make the necessary (major) changes.

The main weakness of this submission is a lack of focus. In modelling the "Cascade Effect", it is not clear what the central contribution of this paper is. At various points, reference is made to crime scripts, cybercrime harms, division of labour, forms of cybercrime (p.2), along with kill chains (p. 8), and also some discussion of definitions/demarcations of cybercrime itself (p. 1). It would be much clearer if the author(s) selected one key area which they felt this paper made a clear and direct contribution to. At present the clearest candidates appear to be crime script analysis (not just Hutchings/Holt but others as well); division of labour/specialization (e.g. Lusthaus, Industry of Anonymity, ch 3); or the linkages between cyber dependent and cyber enabled crime (Furnell, Wall and a number of others). But it is possible, once other changes (below) are made, that another area of contribution appears most relevant to the author(s). In short, what value does a "cascade" model add that is not already covered by other terms/debates in the literature? This is the key element that should be solved and added in, so that this paper is publishable.

The other important point is that the paper sits awkwardly between a deductive and inductive approach. By opening with a cascade model and then seeking it out in other cases, the paper signals a deductive approach, but then pulls away from this in its later stages. Based on the analysis carried out, it also seems that this cascade model is not a good match for many of the cases that were studied, particularly those involving a larger financial motivation. It seems better to adopt a more inductive approach. The author(s) can still open with the cascade model as the first model they have identified, but be more open to developing additional models, as a result of the extra case analysis and what new patterns emerge. If some of these newly analysed cases do match the cascade model, that's great. It then might be interesting to understand what these cases have in common with each other, as this might suggest future cases that are similar in these respects may also match this model. But for the other cases that don't match the model, it would be useful to delineate alternative models, diagram them, and look for similarities between cases grouped under the same models, which again might also predict the models of cybercrime operations in future cases.

Tied to the above point, there seems to be considerable variation in the analysed cases, especially on the fun vs profit motivation. It would be surprising if such diverse cases all did fit under the same model, so seems sensible to allow for this from the outset. There are also many different kinds of "big data" or just "data". Again some data is financial and can be cashed out straight away (e.g. credit card data). But other data might require more elements to monetise it. These differences are likely to lead to quite different cybercriminal operations, and should be accounted for.

For the research design section, and given that WACCO is a very interdisciplinary venue, more detail should be provided on the methods that have been chosen. It can't be assumed that terms like "intermediate-N"

and “crisp sets” will be easily understood by a diversity of readers, with the current level of detail that is provided.

In part III, it is also unclear how the percentages in Tables 1 and 2 were arrived at, which could be explained further. But including these percentages at all seems unnecessary, as the weak yes and no are probably sufficient on their own. With that said, changing the approach to be more inductive should mean these tables are removed entirely.

Finally, there are a large number of smaller points that could be made, and in general the submission is in significant need of both macro and micro edits to make it more focused, cogent and accurate. As there should be significant revision made to this paper, I am not listing these smaller elements, as they may not even appear in the revised version. Hopefully the author(s) also catch the remaining elements through some major editing.

What is very pleasing about this paper is that it engages with legal documents as a form of data. This is a surprisingly under-used approach, which is a shame as there is much richness that can be provided on cybercriminal operations from these sources (e.g. the work of Leukfeldt, Kleemans and others). This analysis of this data is the key value of this submission, and should be made available to the academic community through publication.

Reviewer's confidence: 5 (expert)

Review 2

Time: Jun 07, 18:59

Overall evaluation: 1 (weak accept)

This paper provides an interesting alternative view on how criminals organize to perform cybercrimes (with a particular focus on data crime), but drawing from a mixed method analysis of court documents. The findings are interesting, and this paper has the potential of attracting more research in this space.

I think that this paper could lead to a very interesting discussion at the workshop, and I appreciate the criminology angle that it takes. I think that the paper would be more accessible to the audience of the workshop though if the authors briefly explained their methods and theories early in the paper (for example in the abstract and intro). At the moment the description reads a bit dry, and it is hard for readers from the computer science field (which will be the majority at the workshop) to understand what the paper is about.

It would also be useful if the paper briefly discussed the relation of this model (and related findings) with alternative models of cybercriminal activity, for example "Framing dependencies introduced by underground commoditization" by Thomas et al.

Reviewer's confidence: 3 (medium)

Rebuttal Letter

We are grateful for the time the reviewers took to read our paper and write their comments. We are delighted to read they enjoyed the mixed method (law, criminology and QCA) approach and value the use of original court materials as data. We feel this is one strong contribution the paper makes, especially as it builds on existing work by Turner and Hutchings.

We welcome most of the points made by the reviewers and we are willing to address them, we would like to start with the points we rebut. As the submission was anonymous we took every step to remove references so we are not sure the reviewers were cognizant of the fact that our submission is a follow-up to a IEEE WACCO/S&P paper. We hope that restating this link will addresses Reviewer 1’s comment about the approach taken. We will

also emphasise that whereas the first paper followed a deductive approach, the current submission follows an inductive approach (grounded theory) based upon original and secondary court data. The fact that the model does not fit all cases is actually a finding in and of its own, a point that can be emphasised. While we understand why reviewer 1 feels that such a finding calls for a separate model, we believe the current submission is not the place to do so, not least because of the stringent page limit of IEEE S&P. For this reason we also shortened the methodology section, particularly the part explaining our inductive approach, data gathering and "intermediate-N" configurational comparative method, crisp sets, use of percentages (which we would be happy to drop for this publication if of little explanatory value). We are happy to reintroduce some of the explanations in the paper, thereby accepting comments made by both reviewer 1 and 2, and, to this effect, could we double-check with the editorial board that we can exceed the 10 page limit (currently exceeded by 13 lines).

A second critique made by reviewer 1 is the paper's apparent lack of focus. We believe that the matter will be addressed without major revisions once the links with the IEEE WACCO/S&P paper are restored (and made clearer). The main contributions of the paper are:

(i) the linkages between cyber dependent and cyber enabled crime;

(ii) the inability of current cybercrime legal categories to reflect such linkages which reflects how practitioners interpret it;

(iii) the importance of 'data' to cyber offending. We will rectify any confusion between data and big data - the point by reviewer 1. The concept of data crime seeks to capture how the ready availability of illegally sourced data is fuelling further distributed cybercrimes;

(iv) It is with respect to the potential practical implementation of the modelling that the paper will contribute to the crime scripts literature. But that literature is not the starting point of the paper. Another practical implementation of the research is with respect to sentencing and mitigation, as is discussed in a separate paper pending decision from the legal journal Computer Law and Security Review.

We are keen to make our contribution clear and we welcome the invitation to include further relevant work, such as by Fuller, "Framing dependencies introduced by underground commoditization" by Thomas et al, and other relevant works mentioned by reviewers and encountered in the interim by the authors (e.g. Hunton's cybercrime execution stack).

We hope that you'll find our plan to improve the paper agreeable and hope to be able to contribute to WACCO 2021. (579 words).

Reviews Phase 2

Review 3

Time: Jun 25, 15:03

Overall evaluation: 1 (weak accept)

This paper uses case studies from court records and media reporting to support the modelling of the cascade effect in data crime where data spreads to a wider range of criminals and results in a greater range of harms as a series of tipping points are passed.

The model is potentially useful for encouraging early interventions before data has had a chance to cascade down to a wider range of cybercriminals where more widespread, varied, and harder to manage harm could occur.

This is not my area enough for me to determine whether this is particularly novel but it does seem useful. The argument could be clarified in places but it is generally OK. Some of the flow charts are a bit confusing

and could be better connected to the text. Would a figure at the end with the complete flow chart help? It would be good to see some significant work on the details before the camera ready.

The issue of reproducibility is not really answered within the paper. Will the annotated records be made available to other researchers or more extensive detailing on where all the individual bits of data for the whole set of cases considered came from?

Minor -----

The blinding for review felt overdone and intrusive. It might have been possible to just refer to work in the third person or alternatively just define papers [X] and [Y] on the first occasion and then use them. Even just using "[Blinded for review]" instead of the rather verbose "[[x] Blinded citation to preserve submission anonymity]" would have been a big improvement.

There were a few minor formatting issues, perhaps resulting from an (understandable given the discipline) unfamiliarity with LaTeX. Tables seem to have been included as screenshots from Word rather than natively (admittedly a bit clunky in LaTeX). This results in decreased clarity, red wiggly lines, and formatting marks in Table IV. Similarly the flow charts should be included as vector graphics rather than raster images and there is a random red dot on each one, perhaps the mouse? Some of these could also be tidied up a bit with lines not connecting properly or the potential for boxes to be combined or deduplicated (e.g. multiple "NO cascade" boxes).

The font size of figures should in general match that of surrounding text but it is rather smaller in many instances.

In B.Stage 1: "67.500" probably means "67,500" though that seems implausible, "67\,500" would be better LaTeX as it avoids the ambiguity between "," and "." between locales while ensuring readability.

The formatting of the references looks a bit odd (lack of indenting). Has it been done manually? Half the point of LaTeX is to avoid that pain.

Rebuttal -------- The authors make some useful points in the rebuttal but it doesn't change my overall view of the score. Some major editing is required. There is good work here that should be published, the question is whether it is ready.

Reviewer's confidence: 3 (medium)

Review 4

Time: Jun 28, 16:56

Overall evaluation: -1 (weak reject)

The paper proposes to model the cascade effect between upstream actions and downstream actions in data breach related crimes. The analysis is based on 32 case studies (based on court records and news reports) and provides a descriptive model on the attack stages and tipping points.

Strengths:

Overall, it is interesting to investigate those real-world cases to explain the rationale behind the cascade model and tipping points.

A large number of case studies are considered to construct the model

Weaknesses:

This paper frequently cites the authors' prior papers that also studied these cases. However, these references are not included in the submission (for anonymous purposes). It seems that the authors could have cited those prior works as a "third-person" so that we can read these papers and see what has been improved upon prior works. Currently, it is difficult to see where the improvements are.

About the paper itself, many of the Tables and diagrams are poorly formatted. For example, Table I has red lines under certain words due to Microsoft Word's grammar check. All the diagrams have inconsistent labels on "Yes" and "No" on the flow chart, and do not explain what the "red dot" represents in the caption. Many of the arrows and lines are poorly aligned with the boxes. These tables and diagrams should be further improved.

Methodology-wise, the analysis focuses on specific phases and tipping points separately with example cases to explain the rationale. However, I have trouble understanding the bigger picture. For example, as shown in Table III, case 2, the defendants were involved in multiple phases (stages 1, 2, and 5) but none of these phases have reached a tipping point. It is difficult to imagine how the defendants could jump over different phases without triggering any tipping point nor forming a cascade. If the defendants have skipped some of the stages (e.g., jumping from stage 2 directly to stage 5), should it still be considered as a cascade?

Also, case 5 is quite similar except that case 5 has reached a potential tipping point after phase 3 (y*). I don't quite understand how this is determined.

It would be helpful if the authors can use several cases (at least two) to walk through the entire process to explain all the stages and tipping points to show how these stages and tipping points are labeled. Right now, the individual stages are explained with *different* cases (IV-B), which fails to show the holistic picture for specific cases.

Reviewer's confidence: 4 (high)

Review 5

Time: Jul 02, 11:17

Overall evaluation: 0 (borderline paper)

========== META REVIEW: REQUESTS FOR THE SHEPHERDING PROCESS ==========

The reviewers have agreed that this paper has good elements and should be published following revision. But some major editing needs to be carried out before publication takes place. The author rebuttal presented a plan for making (at least the initial suggested) changes. We suggest that the author(s) set aside some meaningful time to make the original and new changes, and to edit the submission as a whole so that it is publication ready. This editing should also drop the word count to allow for some of the new elements to be added. The reviewers found the anonymisation approach made the paper difficult to read. Hopefully when these citations are added, reading should be smoother, and the distinction from prior work clearer.

While the authors should address each of the reviews in their original form, these are the main points to work on:

1) SHARPEN THE FOCUS: In modelling the "Cascade Effect", it is not clear what the central contribution of this paper is. The rebuttal made clear what the authors wish to highlight as the focus and to better situate which part of the literature the paper contributes to. 2) DEDUCTIVE VS INDUCTIVE APPROACH: The paper sits awkwardly between a deductive and inductive approach. The authors have outlined which approach they wish to follow, and should make sure this is clear in the revised version. 3) MODEL MATCH AND VARIATION: Some of the cases do not appear to match the model. The authors should better explain the nuances involved here. It would be helpful if the authors can use several cases (at least two) illustrate the entire process to explain all the stages and tipping points for a more holistic/bigger picture. 4) METHODS AND DATA: More information needs to be provided on the methods so that it can be understood by readers from a range of backgrounds. Also would be useful to know more about the data, and what can be made available to other researchers. 5) PRESENTATION OF DIAGRAMS: More work needs to go into the selection and presentation of the diagrams to make sure they are directly/clearly linked to the text, and easy to follow.

6) FORMATTING: Along with text editing, and work on the diagrams, reviewers have noted a number of LaTeX formatting issues, which need to be addressed.

Reviewer's confidence: 5 (expert)