Anonymous-1

All remarks here are my own, and do not necessarily represent the views of any other person or organization.
DAAR is a good idea, and should be done. Bambenek's review has more constructive criticism, which makes it a bit more actionable, but both reviews are positive and I agree with those sentiments.

I have one material concern. Specifically, I share Bambenek's concern that data reduction compromises the DAAR claim that incorrect data or false positives in the RBLs are not a big deal. Constructively, consider building a small array of information about each domain, rather than simply presence or absence on a list. For example, note subdomain present, precise domain present, URL containing domain present. Perhaps subdomain contained in URL is an additional one that's useful. When the RBL maintainers make these more precise claims (subdomain, URL, etc.) rather than list an effective second level domain (eSLD), it's on purpose. I would suggest maintaining a hint of that information. Maintaining counts of unique subdomains and URLs on an eSLD is probably too complex. But a small three- or four-bit array for presence of different related parts (subdomain, URL, etc) seems doable.

The following comments are largely about broadening out the explanation of the benefits or justification of DAAR, to explain it in additional ways. Probably, it's all OK as it is. But here are some things for your consideration.

DAAR supports science, in that it supports "structured observations of the empirical world." In general, blacklist curation seems to be science in this sense. Current anti-abuse work can be seen as essentially forensics --- not in the narrow sense of "digital forensics" but in the broader sense of applying scientific methods to provide evidence about a crime (or, in our case, a policy violation, which is conceptually the same thing). Supporting studies is one of the purposes you list, so perhaps the arguments in this paper might help make that link more explicit? This might be background, not foreground, argument. https://urldefense.proofpoint.com/v2/url?u=https-3A__dl.acm.org_citation.cfm-3Fid-3D3171540&d=DwIGaQ&c=FmY1u3PJp6wrcrwll3mSVzgfkbPSS6sJms7xcl4l5cM&r=igGsy6LddHwo34s-Fe8qM3c2yEf4XcKzX0J_L_eSNRE&m=ClHdEEOwDoX2RFgrEcRBLJq8_YhzXV65XoCN6tXXSgA&s=uOYEVRVQcNpYs68smHpQZj6QvT5PJlLxpuvNanZQiQ4&e=(Or https://urldefense.proofpoint.com/v2/url?u=https-3A__tylermoore.utulsa.edu_nspw17.pdf&d=DwIGaQ&c=FmY1u3PJp6wrcrwll3mSVzgfkbPSS6sJms7xcl4l5cM&r=igGsy6LddHwo34s-Fe8qM3c2yEf4XcKzX0J_L_eSNRE&m=ClHdEEOwDoX2RFgrEcRBLJq8_YhzXV65XoCN6tXXSgA&s=CM0QHflyFfpUp_3188jBhfVEVzLnoO-JF0u-7JlvQXU&e= if the ACM link is not free)

Perhaps worth making it explicit that this project's relationship to GDPR? I don't think it's a big problem. But clearing these hurdles explicitly now that GDPR is in force may be wise? The APWG may be a source of advice.

Definition of spam as a threat

See Brunton (2013) for social history of spam (Spam: A Shadow History of the Internet).

Spam is "the use of information technology infrastructure to exploit existing aggregations of human attention." Note that "exploit" is in there. Security violations are exploits. This provides yet another tie to spam as a security problem. All the technical security exploits conducted by spammers as a result of exploiting human attention are well documented by Ranum and Bambenek and the DAAR paper. There's a good review of the Brunton boook here:

https://urldefense.proofpoint.com/v2/url?u=https-3A__lareviewofbooks.org_article_the-2Dtedium-2Dis-2Dthe-2Dmessage-2Dfinn-2Dbruntons-2Dspam-2Da-2Dshadow-2Dhistory-2Dof-2Dthe-2Dinternet_-23-21&d=DwIGaQ&c=FmY1u3PJp6wrcrwll3mSVzgfkbPSS6sJms7xcl4I5cM&r=igGsy6LddHwo34s-Fe8qM3c2yEf4XcKzX0J_L_eSNRE&m=CIHdEEOwDoX2RFgrEcRBLJq8_YhzXV65XoCN6tXXSgA&s=z8cTylXw9LhpMsX0gJGzmFachDgEw536ptiXBAiBNyo&e=

There's some other good social arguments to pile on, perhaps:

`The principal effect of “making spam scientific,” argues Brunton, was that the business of spam was “left […] to the criminals.” `

Thanks for all the good work!

Anonymous-2

In line with the abuse, we need to ensure that Bullet Proof hosting needs to be identified as one of the areas of abuse online. They are black hosting that service without any issues of the hosting organization and are never questioned and represent a large scale of the underground dark web. Most spam and porn etc., support.

These are cloud equivalent support to all the main attacks and heavy users of IPs as well as hidden proxy services. We need a developed framework to approach the real trouble.