

Internet Watch Foundation Developing an Efficient Monitoring Strategy - Phase 2

Executive Summary

Since the earliest days of the IWF a particular focus for the hotline has been on Usenet newsgroups. The experience of the hotline demonstrates that particular newsgroups, albeit a relatively very small proportion of the total number available, have been a regular and significant repository for illegal content, primarily abusive images of children.

At its meeting in November 2001, the Board approved the following recommendation:

The IWF recommends all ISPs serving UK customers not to host newsgroups which the IWF identifies as regularly containing child pornography, the criteria for "regular" being determined by the IWF Board¹.

The work undertaken by DataTalk is designed to assist the Board in determining the criteria for 'regular'. It is essential for the IWF to maintain a reliable, transparent and effective method of reporting levels of illegal activity since this becomes the basis for action against particular parties. The methods used must be demonstrably robust and balance the need to eradicate illegal activity without penalising legitimate use of the Usenet environment.

An objective sampling process is vital if valid projections are to be made about the content of a newsgroup as a whole. In order to improve on the current process DataTalk proposes a systematic sampling process and then builds upon this to develop a robust model for newsgroup estimation.

Geometric Sampling Approach

DataTalk recommends the application of a geometric distribution to the image sampling process. In this case, the variable 'the number of images viewed until a fixed number of illegal images is found' would be measured. This means that groups need only be viewed until an illegal image is found rather than until a

¹ Minutes of IWF Board meeting of 15 November 2001 - <http://www.iwf.org.uk/about/poli.htm>

certain proportion is reached. The maximum number that needs to be viewed would be 200 before the estimate of the overall proportion in the newsgroup is said to be 1% or above. Over three consecutive monitoring sessions of a group, the statistic used to estimate the proportion of illegal content will be the total number of images viewed during the three sessions. The cumulative effect of making consecutive visits would bring the maximum number of images needed before finding an illegal item down to 300 in order for the estimate of the total proportion to be said to be 1% or above.

Four Newsgroup States

Datataalk would recommend the following model:

- Innocent – newsgroups not reported or those that have been cleared following a monitoring sequence
- Suspect/Probation - test monitoring of groups not currently listed or recently de-listed
- Under review
 - Bailed - close monitoring of groups identified as candidates for listing
 - Parole – close monitoring of 'sentenced' groups to check whether they stay listed
- Sentenced – group identified as containing illegal content and recommended for removal by ISPs. Sentenced groups would remain de-listed for a period agreed under IWF policy, currently 3 months.

Newsgroups change state based upon the outcome of each group of three monitoring sessions (see Appendix 1).

The geometric approach means that the same process and criteria can be applied to groups at any of these stages, which simplifies the model. Operators need not be aware of the current status of any newsgroup being monitored. The monitoring process is identical at all stages, thus helping to eliminate any pre-conceived ideas that may inadvertently colour the monitoring scheme.

In addition to the monitoring criteria laid out DataTalk suggests that the additional criteria for remaining listed is maintained, namely if a single image is found on each of six consecutive monitoring sessions.

Proposed Monitoring Process

Proposed Monitoring Process Flow:

- Newsgroups can be in four states: Innocent, Suspect, Under Review, Sentenced.
- The decision about the current state of a newsgroup is made after each group of three monitoring periods separated by at least two weeks.
- Each monitoring period consists of viewing up to 200 images selected according to the recommended sampling scheme. The period ends when the first notifiable image is viewed is then reported and the number of images viewed is recorded
- Newsgroups move up the states if less than 300 images were viewed over the three monitoring periods.
- Newsgroups move down the states if no illegal images were detected over the three monitoring periods, i.e. 600 images were viewed.
- If less than 200 images are available during any given monitoring period, the sessions are combined and estimates are based upon the detection across either 2 sessions or a single session as shown in the tables in Appendix 3

There is currently a difficulty in making a subjective determination as to whether or not the presence of, for example, 0.97% or 1.01% of illegal material within a group should be captured under the 1% rule. The geometric approach to sampling enables DataTalk to provide definitive tables on the basis of which an objective and defensible decision could be made in such cases.

Policy and personnel considerations of Geometric Sampling:

Whilst DataTalk recognises that these issues are outside the scope of this report, they have arisen in the course of the project and are flagged here for consideration.

- Human resources:
Moving to the geometric approach would reduce the intensity of activity required, provide an enhanced sense of task significance, and minimise the number of abusive images which analysts will have to view in the course of monitoring. These factors may have a significant effect on the retention of trained and skilled staff who are essential to the effectiveness of the hotline.

- Public relations:

Given that each individual image constitutes evidence of a criminal act against a child, the 1% threshold may be perceived as conveying an implication that the presence of illegal content at a level below the threshold is acceptable. With this in mind we recommend continuing the policy of continuing to de-list any newsgroup if an illegal image has been found on each of six consecutive occasions.

- Responsibility to ISPs:

By using the geometric approach involving the identification of just one illegal image, Internet content analysts will be able to action all images found within the monitoring activity.

Monitoring the effectiveness of IWF activity:

It is important to estimate the likelihood of Internet users accessing potentially illegal images, in order to assess the level of protection that IWF's activities are affording to the public. Using capture-recapture methods drawn from biological population studies IWF would be able to estimate illegal volume and assess the effectiveness of its activity, based on the number of previously actioned reports received.

Conclusion:

Objective sampling methods and a geometric approach provide the basis for robust estimates of illegal content. Primarily this approach has been discussed in the context of newsgroup monitoring, but it could be applied equally to other 'user group' areas of the Internet such as online clubs and communities.

In addition, capture-recapture estimation of population size and IWF penetration provide an opportunity for balanced estimation across a whole variety of sources and times, supporting the comparison of activity in an objective fashion. This approach can be extended from assessment of the overall annual statistics to the day to day management of resources.

Appendix 1

