The TalkTalk 'BrightFeed' System

Richard Clayton, March 2011

Introduction

On Tuesday 11th January 2011 I put on my 'hat' as a member of the Advisory Council of the Open Rights Group (ORG) and accompanied Jim Killock, ORG's Director to a meeting with TalkTalk (the large UK ISP) to learn about 'BrightFeed', which is the working name of their new online content control system (they previously called it 'NetworkSecurity' internally).

We received a wide-ranging briefing about their system, not only from the point of view of a user, but we were also given a lot of technical details about how it will work.

The meeting was entirely on the record, and they fully understood that we would be writing about what we learnt from the meeting. I also provided an initial draft of this document to TalkTalk and they pointed out some inaccuracies that I have now corrected.

I have tried to stay away from any commentary as to the social, moral or legal issues that might arise from the system, but have concentrated on the technicalities of its operation.

Nothing within these notes should be taken as any sort of approval or endorsement, or indeed disapproval or criticism, of TalkTalk's system design or implementation. This document is merely a record of my current understanding of how their system works.

As a road-map to what's in these notes, I shall be discussing:

- TalkTalk's current content filtering systems;
- the operation of the BrightFeed system;
- how BrightFeed assesses websites visited by TalkTalk customers;
- and, a brief mention of regulatory issues.

A TalkTalk's current content filtering systems

1. TalkTalk currently operate, or offer to their users, a number of content filtering systems.

2. TalkTalk run a blocking system that will prevent access to the websites on the Internet Watch Foundation (IWF)[1] blocking list. The IWF maintains a list of URLs (several hundred at the time of writing) which contain child sexual abuse images. Almost all UK ISPs have some type of blocking system that uses this list.

3. The IWF-list-based system is completely separate from the other systems described in this document. It will continue to operate exactly as at present even when new systems are deployed.

[1] `http://www.internetwatchfoundation.org.uk`

4. TalkTalk filter email spam as it arrives at their mail servers, and also scan this incoming email for 'malware' (malicious software, sometimes called 'viruses', 'worms' or 'trojans').

5. Malware can pass confidential information to criminals (so that they can steal from bank accounts). Malware often causes the computer to become part of a 'botnet' sending email spam or fraudulently clicking on adverts. Other types of malware will interfere with browsing so as to display unwanted adverts or redirect web page views in such a way as to defraud legitimate affiliates who get paid for customer referrals.

6. TalkTalk provides 'Super Secure Boost' whose features are provided by F-Secure. This is an anti-malware program that runs on end-user computers. Use of this product is optional.

7. TalkTalk also provides 'Magic Desktop' an application aimed at the parents of younger children. Along with a range of educational games, there are some parental controls and 'my first browser'. Use of this product is also optional.

B The 'BrightFeed' blocking system

8. TalkTalk's new system ('BrightFeed') operates at the network level. It affects all traffic going to or from a particular customer. It is not possible to arrange for it to affect only some computers (children's computers, for example) and have no effect on the rest of the household. However, this does mean that, unlike existing systems, it will apply the same set of content controls to non-PC devices, such as games consoles.

9. The system uses equipment from Huawei-Symantec, and in particular the parental controls it uses are similar to those in other Symantec products.

10. The system only operates on HTTP traffic over the standard TCP port 80. HTTP traffic over other ports is not affected. Traffic over port 80 which is not HTTP will be ignored. Encrypted (HTTPS) traffic will also be ignored (whether over port 80 or the usual port 443) because the system will be unable to decrypt it.

11. When an HTTP access is made on port 80 the BrightFeed system will pick out the URL (the site to be visited) and will look it up in a TalkTalk hosted database. If the site is 'bad' a 404 error code will be returned and the user will see a web page that explains that blocking has taken place.

12. If the site is 'good' then the request will be passed to the remote site, which will return whatever it wishes in the normal way. The remote site will be unaware that any check has occurred.

13. The system has a number of different 'bad' categories, such as 'Pornography', 'Mature content', 'Lingerie', 'Alcohol', 'Computer Hacking', 'Cult' and so on. The account owner can personalise which categories will be active for their connection – so they could perhaps enable 'Politics' and 'Sex Education' if they believed this was suitable for their family's use of the Internet.

14. An important 'bad' category is the presence of malware that is capable of infecting a visiting computer. TalkTalk believe that visiting websites is the major way through which their customers are being compromised by malware.

15. If customers do not opt-in to the BrightFeed system then no blocking will occur. The filtering (the blocking of 'bad' sites) is only performed for customers who have chosen for this to happen and when the website falls into a category they have chosen to have blocked, i.e. this is an 'opt-in' system.

16. No customer specific logs are created for blocking events.

17. The 404 blocking page will allow someone in possession of the relevant password to disable the blocking.

18. There is provision for setting the time periods when Internet access is blocked or when it has a different list of categories active (preventing access to social network sites during a 'homework' period for example).

C BrightFeed's collecting of URLs

19. There is another aspect to the BrightFeed system that applies to all TalkTalk customers, whether they have opted-in or not. The mechanism which utilises the URLs from HTTP requests on port 80 is active for every customer, and all URLs will be passed to the database system, even when the customer has not opted in the BrightFeed blocking mechanism.

20. If the database system was previously unaware of the URL then it will create a record of this new URL. Only the URL is stored. No record is kept of which customer was attempting to access the URL.

21. The new URL is passed to a scanning system whose task is to assess whether or not the URL contains malware. The system adds the URL to the list it maintains of sites that need to be checked.

22. After a time, which TalkTalk intend to be a short time, an automated access to the URL will be made by a TalkTalk machine. This machine will attempt to determine whether the URL is malicious, in the sense of attempting to infect a visitor with malware.

23. The automated access will take no notice of any `robots.txt` file. However, it has a characteristic 'user agent' which will identify the scanning system.

24. Whether or not a website contains malware and should therefore be labelled as 'bad' will be a judgement made by 'anti-virus' software provided by Symantec.

25. If the website is malicious it will immediately be marked as 'bad' in the TalkTalk database. From that time onward, visitors to that URL (who have opted in to the BrightFeed system) will be protected. Unavoidably, any TalkTalk customers who visit the website before the assessment takes place will not have had their access blocked.

26. If the site is not malicious, knowledge of this 'good' URL will be retained for a period to record that the scan for malware came back clean. This avoids any need to repeat the scan if others visit the same URL.

27. After a period, URLs are discarded from the database so that a 'good' or 'bad' result will not be recorded forever.

28. The Huawei-Symantec system is deployed world-wide and 'bad' URLs are learnt about from central systems, but 'good' URLs do not come from central systems but are only learnt about locally.

29. URLs that are labelled 'good' or that have yet to be scanned will always stay within the confines of the automated system within the TalkTalk network and will never be inspected by any human.

30. However, URLs that are deemed to be 'bad' will leave the automated system and be reported to TalkTalk personnel. They may be used, for example, to allow TalkTalk to work with legitimate website owners to help resolve whatever malware infection they may have.

31. No URLs, whether 'good' or 'bad' will ever leave the TalkTalk network and they are never passed to the Huawei-Symantec central systems.

32. URLs consist of several parts, the familiar `http://hostname/path` string can be preceded by a *userinfo* section (`user@password`) although few browsers now support this for `http` and `https`. More common is the addition of a *query* string (`?parameters`) and/or a *fragment* (`#label`). Clearly some URLs will be specific to a single individual. If a URL is user-specific then, besides any issues relating to personal data, assessing it as 'good' or 'bad' would not be of general benefit.

33. We failed to discuss removal of the userinfo, query and fragment during the meeting with TalkTalk. In subsequent correspondence they have said "We strip URLs of personal information" but that the "exact terms we strip have not been disclosed".

D Regulatory oddments

34. TalkTalk do not expect to change their Terms and Conditions for their Internet access products when the new system is deployed, although of course they reserve the right to make changes in the future as permitted by those Terms and Conditions. It is their belief that the operation of the BrightFeed system comes fully within the existing Terms and Conditions.

35. At the time of writing, the system is being trialled but is not at present being offered to customers as a service.

36. The system has been tested on a trial basis in the past. Those earlier trials were not made public for reasons of commercial confidentiality. However, TalkTalk told us that having subsequently had discussions with the Information Commissioner's Office (ICO) they can understand the benefit of informing the ICO, and their customers, of the nature of trials of this type of system.

Dr Richard Clayton
Cambridge, UK
March 2011