REMOTE INTERNET OBSERVATION

James Sutherland, Natalie Coull, Ian Ferguson
Security Research Group, University of Abertay Dundee

UNIVERSITY of ABERTAY DUNDEE

The Scottish Institute for Policing Research

SRG Security Research Group

RIO – Remote Internet Oversight

The RIO project was driven by two key developments. Internet capable devices other than PCs have proliferated recently: smartphones, "smart"TVs, games consoles and even central heating thermostats can now be Internet-connected. Meanwhile, several courts have rejected outright prohibitions on Internet and computer access as being unduly restrictive – so an effective way of monitoring Internet usage by offenders is increasingly important.

To address this need, the University of Abertay's Security Research Group is developing a remote monitoring solution, RIO: a modified Internet router which can be installed in the subject's home and efficiently monitor Internet usage, reporting in real time to a central server for later analysis.

Duplicating all the content downloaded is not an option, for multiple reasons. On a technical level, the bandwidth required would be prohibitive, since typical home connections can download data up to ten times as quickly as they can upload it. Legally, it would present three distinct problems: liability, in case private content should leak; copyright; and should the subject of monitoring download any illegal content, we would now also be guilty of duplicating and possessing it.

Instead, for web access the monitoring device sends the address of each object requested, along with a cryptographic hash of that object – a numerical fingerprint of that data – and keeps a local full copy. The hashes can be checked against both whitelists of known safe content (such as routine system updates) to avoid wasting resources, and blacklists of inappropriate content (such as CEOP's database of known child abuse images).

Encrypted connections need special handling. For ordinary services, we intercept and impersonate the server, so encryption is no barrier to monitoring; for Extended Validation sites (commonly used by banks and other high-sensitivity services) we simply log the time, destination and volume of traffic.

Filtering versus monitoring

One approach is to apply a filter, preventing access to material which is known beforehand to be prohibited. This, however, has significant failings:

1. Limited effectiveness and redundancy
Any such filter would be limited to material already known to be illegal. Of course, this cannot possibly be a comprehensive list, meaning some material would go unfiltered — moreover, such a filter is already in place through the IWF (Internet Watch Foundation), which maintains an extensive list of known locations of illegal material for use by ISPs (Internet Service Providers) through filtering systems such as CleanFeed. Equally, any filter will sometimes block material it shouldn't — CleanFeed notoriously restricted access to much of Wikipedia for four days in December of 2008, as well as WordPress.com and The Wayback Machine at various later dates.

2. Potential for misuse
Installing any such filtering system would also imply installing a list of known illegal sites. As Richard Clayton of Cambridge University demonstrated in 2005[1], it is quite feasible to extract the list of filtered sites from the filtering system — so this approach would provide sex offenders with a convenient in-house list of all web servers known to contain child abuse material. It would also reveal which sites are not yet known to authorities.

3. False sense of security
The presence of a filter intended to prevent access to illegal material would risk giving everyone involved a false sense of security, as if all such access were impossible rather than merely impeded.

Conversely, the approach used in RIO — allow but monitor all downloads, storing a hash value for later analysis — means that if known illegal material is republished in fresh locations, that access will still be detected and recorded, alerting operators to the new location.

Hashing

One important technique in the design of RIO is a mathematical 'hash', a numerical fingerprint of a block of data. This generates a consistent number of a fixed size from any input — for example, the word 'fish' has a fingerprint of 64875fcccaac069fcb3e0e201e7d5b9166641608. Just like a fingerprint, this is close enough to being unique to be used for identification purposes; also like a fingerprint, it does not give any information about the original data — there is no way of converting that number back into 'fish' besides simple brute force, trying words until you find one that matches. (This example uses a SHA1 hash, with 2^{160} possible values, so any 'collision' — a false positive match, where two different items gave the same fingerprint — would be virtually impossible to find even with substantial resources devoted to the search.)

This is extremely useful in identifying known illegal images, before or after the fact: by recording this numerical fingerprint for each file downloaded, it's a simple job to check this list of files against any list of proscribed material, without needing to retain or copy the material itself. When new illegal content is identified, it can be added to the list and re-checked against older download logs, detecting access to the illegal material even prior to its identification as illegal.

References

1. http://www.cl.cam.ac.uk/~rnc1/cleanfeed.pdf
Phone image: www.webdesignhot.com