WOMBIR Jan 2021: Position Paper

Richard Clayton, Cambridge Cybercrime Centre, University of Cambridge

What we do

The Cambridge Cybercrime Centre is a multi-disciplinary initiative combining expertise from the University of Cambridge's Department of Computer Science and Technology, Institute of Criminology and Faculty of Law. We started work on 1 October 2015. We collect and collate datasets relating to cybercrime and share those datasets under license with other academics. This enables them to work with the data, using their skills to better understand cybercrime, without the need to collect the data themselves.

Some of our datasets are collections of posts from underground forums and open chat channels – but others are of network activity: we have been tracking reflected amplified DDoS attacks since 2014, we operate honeypots that aim to collect IoT malware (242 000 unique samples and counting) and we track 'Mirai style' scanners which use distinctive SYN packets to rapidly scan the Internet for devices which are candidates for brute force attacks.

We share this data, but not in the usual way after paper publication (and hence months or years later). We share data as soon as it is collected, and sometimes we have not even looked at it ourselves.

How we do some of it

However, we do have data we cannot share. We have installed wiretaps on the University's JANET links and are able to monitor all incoming and outgoing traffic to our three /16s (and a bit). This is legal in our jurisdiction because we are working alongside, and with the full permission of, the people who keep our network safe. Our ethical case only allows us to inspect traffic flows which are initiated from outside the University and we may never look at email traffic. This means that we can look at traffic relating to scans and attempts to compromise University machines, but we cannot look at whether any devices are beaconing out.

The reason that we cannot share this data is legal – the statutory defence we have against a charge of unlawful interception (wiretapping) cannot be extended to third parties. We can (and do in the case of the Mirai scanning data) extract particular datasets from the mass of traffic, and those we do feel able to share.

What might we do more ?

We would like to share more data ... and my reason for spending time engaging with the WOMBIR workshop is to understand:

- are there other derivative datasets that we could extract and make available to the community ?
- is there any mileage in constructing synthetic datasets which can be freely shared which are indistinguishable, statistically, from real traffic being seen at Cambridge ?
- are there valuable collaborations in prospect where we operated security experiments (for our legal carveout is solely based on making the Cambridge network safer) on our systems and on our data, but the techniques that were being tried out were not thought up by us ?

Sharing wiretap data is always going to be very challenging. Another approach which might be worth exploring is joint work with other institutions who have (or could lawfully build) similar wiretapping systems. We might not be able to share any data at all, but we could learn a very great deal from comparing what we saw... one can do natural experiments within a diverse institution (do the biologists get more hacking attacks than the chemists because their IP is currently more valuable ?) but the results are more interesting if the apply in other places as well.[1]

I look forward to interesting discussions.

[1] The Cambridge History Department has more security concerns than you might suspect, because some of senior ex government security agency people work there, and attackers might want to rifle through their papers or read an advance copy of a new book!