On the Origins of a Thesis

Steven J. Murdoch[1]

University of Cambridge
Computer Laboratory
Cambridge, United Kingdom

Abstract

A PhD thesis typically reads as an idealised narrative: how would the author perform their research had the results and conclusions been known in advance. This rarely occurs in practice. Failed experiments, unexpected results, and new collaborations frequently change the course of research. This paper describes the course of my thesis, and how its initial topic of distributed databases changed to covert channels, then anonymity, before eventually settling on links between the two. This illustrates concrete benefits from informal interactions, low-overhead collaboration, and flexibility of research project plans.

For my PhD thesis, I was originally going to study distributed databases and their security, but what I ended up covering was anonymous communication networks – systems to protect users' online privacy by hiding their communication patterns – and how to improve their security by drawing ideas from the field of covert channels. This substantial change was made possible by the flexibility offered by the University of Cambridge, and assisted by collaborations built up at conferences and facilitated online. My resulting thesis was awarded the ERCIM Security and Trust Management Working Group prize, and this paper presents a more formalised version of the talk I gave when accepting that prize.

There is clearly a significant gap between what I planned to study and how I finished. To understand how this happened, it is helpful to look at how I started in October 2002. My supervisor Markus Kuhn encouraged me to develop other small research projects, especially at the start of my studies. Some of these led to publications or software, a few eventually became the core of my thesis, and many produced nothing except useful experience. Whilst I had expectations as to where each of the projects would lead, more often than not I was wrong.

In my first year I worked on finding security weaknesses in the RFID access control system used in my building, studying Security Enhanced Linux, finding weaknesses in a proposed banking security protocol, building a model checker with an integrated Lua-based extension language, writing a survey of markup languages, integrating a one-time password system into the Linux authentication subsystem,

[1] http://www.cl.cam.ac.uk/users/sjm217/

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: www.elsevier.nl/locate/entcs

and generated *n*-gram statistics from the British National Corpus. All were interesting projects, and my banking security thread of research has continued to this day, but if I were asked to find a starting point for my eventual thesis, it would be when I decided to win an Xbox.

1 Local covert channels in games

In December 2002, the Cambridge University Computing Society announced a programming competition [1], for which students were to submit programs to play Connect-4. Each program would play all other entered programs and the winner of each match would be awarded two points or, in the case of a draw, both programs would be given one point. The team who entered the winning program would be awarded an Xbox donated by Microsoft.

My colleagues from the Security Group, Stephen Lewis, Piotr Zieliński, and I designed and implemented a set of programs for the competition. The variant of Connect-4 selected by the competition organisers, where players could pass, was impossible to win because it was simple for either player to force a draw. Our insight was that our programs didn't have to play better Connect-4 to win the Xbox – we just had to win more matches. So our programs would normally force a draw, but also signalled their identity to the opponent over a hidden communication mechanism, termed a covert channel. If one of our programs detected that its opponent was a specially designated one from our group, it would deliberately lose the match, hence giving one of our programs the winning score with a safe margin.

In addition to winning the Xbox, we generalised the work and presented the results at the 2004 Information Hiding Workshop [11]. The workshop was hosted in Toronto, and co-located with Privacy Enhancing Technologies (PET), so I attended both. While at PET, I met many members of the anonymity community, including the designers of the Tor anonymous communication network [2]. On returning to Cambridge I became more involved with the Tor project, contributed to the documentation, and operated one of the network's servers.

2 Low-cost traffic analysis of Tor

The Tor network gives its users anonymity by routing their communications over multiple servers, hiding who is communicating with whom and what path data takes through the network. Another of my colleagues at the Security Group, George Danezis, suggested some potential vulnerabilities in Tor, based around the expectation that the latency of a connection to a server will depend on its load. As I was already operating a Tor server, we were in a position to test this hypothesis.

We found that indeed there was a noticeable increase in connection latency when we injected extra traffic through a given server. An attacker who wants to find out who is accessing a web server he controls could leverage this phenomenon by injecting a characteristic load pattern into a stream he wants to track, and simultaneously look for the same pattern in the latency of all Tor servers. If a match is found he has good assurance that the target stream is passing through that server, and may use this information to de-anonymize the end-user.

Our paper on this topic was accepted to the IEEE Symposium on Security & Privacy (Oakland 2005) [7].

3 Embedding covert channels into TCP/IP

In early 2004 I also investigated the newly deployed anti-counterfeiting measures in popular image manipulation packages such as Adobe Photoshop. The software detected if an image being processed was of currency and refused to open it. I presented the initial results of this project at the 2004 Information Hiding Workshop, which drew the interest of Ben Laurie. Together we continued to reverse-engineer the digital watermark detection software, and I presented our results at the 2004 IT security summer school run by Maximillian Dornseif in RWTH Aachen.

At the summer school I worked on a wide variety of computer security projects, and I was also encouraged to attend the 21st Chaos Communication Congress (21C3), held in December that year. There, I presented my work on reverse engineering the anti-counterfeiting software [9], and also the results of one of my projects at the summerschool – detecting information inadvertently left in the EXIF headers of JPEG images [8]. While at 21C3, I attended a talk by Joanna Rutkowska on the TCP/IP steganography scheme she developed – Nushu [12]. Due to my attendance at the Information Hiding Workshop, I was familiar with the steganography literature and noticed some potential flaws.

On our return to Cambridge, Stephen Lewis and I studied Nushu and other TCP/IP steganography schemes and found weaknesses in them all, so wrote a paper for the 2005 Information Hiding Workshop [10]. One of the most powerful attacks we developed was to characterise the TCP/IP parameter that is being used as cover, such as the initial sequence number (ISN), then look for deviations from this fingerprint in network traces where steganography is suspected.

4 Temperature-based channels

The TCP/IP ISN not only shows the operating system of its target, but also indicates its clock skew – how its real time clock deviates from true time. This is a consequence of Linux using a 1 MHz timestamp as part of the ISN. I discovered this effect while writing software to implement the steganography detection algorithms, and sought to explain an unexpected false-positive (which turned out to be due to clock-skew between my computer and the target host).

One of the talks presented at Oakland 2005 was on using clock-skew to deanonymise hosts [3]. The authors showed that a computer's clock skew is stable over time, but varies across different computers even of the same model. Through the TCP/IP timestamp option of collected packets, normally a 1 kHz clock, it is possible to estimate remote clock skew and detect whether two packet streams were generated by the same host.

As I found that a 1 MHz counter could be extracted from the Linux ISN, I expected that using this for clock skew estimation would give a substantial reduction in measurement error over the 1 kHz TCP timestamp option. However, when testing this I found an unexpected result – the error was not smaller, and on further investigation the size of the error changed during the course of my overnight experiment. I eventually realised that the host I was measuring ran housekeeping tasks at 1 am, which warmed up the CPU and changed the clock skew sufficiently for my measurements to detect.

The phenomenom was effectively a covert channel, similar to the one used in my previous paper on Tor, but applicable even if that attack was resisted by preventing different streams affecting the latency of each other – one of the defences I proposed in the paper. Following the methodology of my previous paper I could therefore use clock skew to measure temperature, which in turn was affected by CPU load and hence detect load patterns injected. I experimentally verified this attack against Tor and presented my paper on this topic at the 2006 ACM Conference of Computer Security (CCS) [5].

5 Covert channel vulnerabilities in anonymity systems

Throughout this period I had been continually revising the scope of my PhD thesis. I had developed some software implementing prototypes for aspects of the distributed database system which formed my original proposal, and presented the plan [4]. Following my growing interest in anonymous communications I also intended to integrate the transport protocol of my distributed database design into Tor, to improve the performance of web page download.

However, the bulk of my work was on anonymity and in particular covert channels. Through my study of Tor in the preceding years, it became increasingly clear that covert channels were a useful, and under-examined, technique for discovering and modelling attacks on anonymity systems. I therefore chose to write my thesis on this topic by formalising the relationship between anonymity and covert channels and extending the papers that I have discussed above [6].

6 Conclusion

It is difficult to fully capture the links between the activities, research, and collaborations which eventually led to my thesis, and I am sure there are many which I do not realise, but a few of the more important ones are shown in Figure 1.

It was through attempting to win the Xbox that I came to write the paper on using covert channels to win a Connect-4 competition. The paper which I wrote on this topic, at the Information Hiding workshop both informed me about steganography and encouraged me to develop my watermarking studies. Through presenting my watermarking results at 21C3 I became further interested in TCP/IP steganography. Following from my attendance at the Information Hiding Workshop I was also introduced to the Privacy Enhancing Technologies community and became involved in Tor and came to write my paper on traffic analysis. Finally, by attending the talk on clock skew at Oakland 2005, when combined with my experience with TCP/IP steganography, I was able to write my paper on temperature covert channels, and eventually my thesis.

The surprising links which contributed to my thesis, many of which were unplanned, illustrates the difficulty in predicting the path of research. It also shows

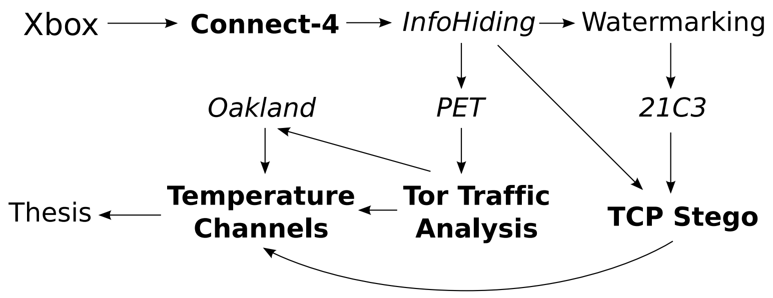

Fig. 1. Links between my thesis chapters (bold), conferences (italics) and topics

the value of collaboration – without the motivation and skills of the other participants the end-result would have been impossible. Though lightweight communication and distributed development practices I was able to rapidly develop shared projects, avoiding administrative overhead which would take time, resources, and sap enthusiasm. However, while the collaborations were often remote, they were forged during person-to-person meetings, showing a further important, but hard to measure, benefit of academic conferences.

References

[1] Cambridge University Computing Society, *Winter programming competition* (2002), http://www.cucs.ucam.org/competition.html.

[2] Dingledine, R., N. Mathewson and P. F. Syverson, *Tor: The second-generation onion router*, in: *Proceedings of the 13th USENIX Security Symposium* (2004).

[3] Kohno, T., A. Broido and k. claffy, *Remote physical device fingerprinting*, in: *IEEE Symposium on Security and Privacy* (2005), pp. 211–225.

[4] Kuhn, M. G., S. J. Murdoch and P. Zieliński, *Compounds: A next-generation hierarchical data model*, Microsoft Research Academic Days, Dublin, Ireland (2004), poster presentation.

[5] Murdoch, S. J., *Hot or not: Revealing hidden services by their clock skew*, in: *CCS '06: Proceedings of the 9th ACM Conference on Computer and Communications Security* (2006), pp. 27–36.

[6] Murdoch, S. J., "Covert channel vulnerabilities in anonymity systems," Ph.D. thesis, University of Cambridge (2007), technical report UCAM-CL-TR-706.

[7] Murdoch, S. J. and G. Danezis, *Low-cost traffic analysis of Tor*, in: *IEEE Symposium on Security and Privacy* (2005), pp. 183–195.

[8] Murdoch, S. J. and M. Dornseif, *Hidden data in Internet published documents*, in: *21st Chaos Communication Congress*, Chaos Computer Club e.V., Berlin, Germany, 2004.

[9] Murdoch, S. J. and B. Laurie, *The convergence of anti-counterfeiting and computer security*, in: *21st Chaos Communication Congress*, Chaos Computer Club e.V., Berlin, Germany, 2004.

[10] Murdoch, S. J. and S. Lewis, *Embedding covert channels into TCP/IP*, in: M. Barni, J. Herrera-Joancomartí, S. Katzenbeisser and F. Pérez-González, editors, *Information Hiding Workshop (IH 2005)*, LNCS **3727** (2005), pp. 247–261.

[11] Murdoch, S. J. and P. Zieliński, *Covert channels for collusion in online computer games*, in: J. Fridrich, editor, *Information Hiding Workshop (IH 2004)*, LNCS **3200** (2004), pp. 355–369.

[12] Rutkowska, J., *The implementation of passive covert channels in the Linux kernel*, in: *21st Chaos Communication Congress*, Chaos Computer Club e.V., Berlin, Germany, 2004, http://www.ccc.de/congress/2004/fahrplan/event/176.en.html.