From: David Kotz
To: Henning Schulzrinne
Cc: Harry Mussman; tristan@cs.st-andrews.ac.uk; "Henry Yeh"; "Helen Nissenbaum"; "Aaron Falk"; celliott@bbn.com
Subject: Re: GENI project interested in CRAWDAD experiences kotz r1
Date: Monday, March 30, 2009 12:26:07 AM
Attachments: CRAWDAD-Contribution-Agreement.doc
CRAWDAD-Contribution-Agreement-2.doc

On Mar 28, 2009, at 11:12 PM, Henning Schulzrinne wrote:

> Certainly, a write-up of the experiences would be great. In addition, we can also do an interview by phone at some point, maybe based on your short write-up.
>
> Henning
> (wearing my GENI Opt-In co-chair hat)

Colleagues,

Here are some quick thoughts I have on our experiences with CRAWDAD, including our efforts to collect wireless traces at Dartmouth, and our efforts to obtain traces collected by others, to share through CRAWDAD. I hope this helps.

Collecting Wi-Fi traces at Dartmouth:
- Our data collection was never opt-in, nor even opt-out. Indeed, we have operated with a waiver of 'informed consent' from the IRB, because it is simply impractical to notify, let alone obtain consent from, all users on an open wireless network. [Dartmouth has only recently begun to require authentication for its wireless network (and soon, wired network). Although the authentication mechanism may provide some opportunity to provide notice or obtain consent, the network also remains open to the public without any splash screen. Eventually that might change.]

- We collected our first traces in 2001. Over time our techniques have become more aggressive -- we now sniff 802.11 frames from 200+ radios across campus rather than ethernet frames from the wired side of a few dozen APs -- and we have attracted more oversight on campus. It has taken over two years to obtain permission to install and operate the 200 new sniffers; we hope to go live in a few weeks. The challenges arose in part because of the community's increased awareness of the legal issues related to network-trace collection, notably, Sicker, Ohm, and Grunwald's IMC07 paper "Legal issues surrounding monitoring during network research". We conducted extensive security and privacy planning, narrowed our collection to MAC headers only, developed a secure trace-collection infrastructure, prepared a robust inline anonymization process, and drew up some careful policies of operation. We retained outside consultants (one legal expert and one technical expert) to review our plans and provide advice. We collaborate closely with the network-operations team. From the perspective of research value, we may have compromised too much-- I often wish we could have IP and TCP headers, for example. Ultimately, the community (perhaps led by GENI?) needs to develop a set of community norms, and push for clarified laws and regulations. Both Dirk Grunwald and Mark Allman have been publicly pushing for the development of a community effort.

- I happen to have some text (the appendix to a forthcoming survey on anonymization methods) that describes our effort to set up our current wireless sniffer infrastructure at Dartmouth; I append that 'case study' below.

- As an aside, in my conversations with lawyers during that process, I understood that life would be much simpler if we had been able to obtain informed consent (whether opt-in or opt-out) from the research subjects. Keep in mind that some states, like NH, requires consent from both (all) parties to a conversation in order to record a conversation, so it may be impossible to obtain informed consent! Although primarily intended for telephony, I think the same rules apply to network connections. Disclaimer: I am not a lawyer.

Sharing traces others have collected, through CRAWDAD:
- People usually come to us with traces already in hand - we rarely provide advice about how to collect traces in the first place. Furthermore, we only ask that they certify that they have complied with the necessary laws and, where appropriate, IRB approvals, to collect their traces. Thus, we have little knowledge of, or insight into, the opt-in, opt-out, or notification methods they may have used during trace collection.

- For traces whose CRAWDAD release will be their first public appearance, we require the contributor (or their legal office) to sign a legal agreement with Dartmouth College. I attach our agreement form (we have two versions). This requirement has lost us many exciting traces, because some organizations (particularly large corporations with deep pockets) have lawyers that don't like to sign agreements assigning them the liability for mistakes. Nonetheless, it is important that we protect CRAWDAD, and Dartmouth, from any liability resulting from the collector's failure to comply with laws or norms regarding trace collection, or failure to anonymize the data sufficiently.

- For traces that are already publicly available on the open web, we simply ask the contributor to send an email approval for us to mirror the trace on the net. We keep a printed copy of their email and of the website where the data appeared, for our records. We don't require a signed agreement in this case, because they released the data to the world, not us.

- We require all users of the traces to sign a click-through agreement before they can obtain a CRAWDAD account and download traces. You can see the text on http://crawdad.org/registration.php .   I'm not sure how many users actually read this text, but the gist is as follows:
    - do not redistribute the traces (we want control so all users see the agreement, and so we can collect statistics)
    - respect the privacy of the human subjects -- do not try to reverse the anonymization!
    - acknowledge CRAWDAD in your publications, please
    - we can share download statistics with the contributors
    - Dartmouth doesn't make any promises about the data, and disclaims any liability

Case study: Dartmouth Internet Security Testbed.

It is often tremendously difficult to obtain permission to collect network traces on a production network, not to mention the logistical and technical challenges of establishing a robust and effective trace-capture system. In this appendix, we offer as a case study our experiences in deploying the Dartmouth Internet Security Testbed (DIST)~\cite{web:dist}. We hope this case study offers practical lessons for others who may wish to collect network traces within their own enterprise.

Two years ago, in January 2006, we sought to build a large testbed for conducting network-security research at Dartmouth College. The testbed would contain both wired- and wireless-network components, and would cover a substantial fraction of the campus production network. The wireless-network infrastructure would be used initially for trace capture, but the hardware would also be useful for other wireless-network experiments including controlled studies of Wi-Fi network attacks. The wired-network infrastructure would only be used for capture and real-time analysis of traffic on the campus backbone network. Although research was the primary purpose of and motivation for the infrastructure, the Dartmouth network-operations group was enthusiastically interested in leveraging the infrastructure and the researchers' results for operational monitoring of the network.

The wireless-network infrastructure consists of about 220 Wi-Fi access points, of the same brand and model Dartmouth uses to provide its production Wi-Fi network on campus. We had developed a scalable network-monitoring and intrusion-detection software base in our MAP project~\cite{sheng:map}, in which we reflash the Aruba AP70 access points with OpenWRT Linux and run our own software for sniffing on the WI-Fi network interface. This software uses pcap to capture Wi-Fi frames and packs multiple frames into a custom format for transmission to our central server for real-time analysis and (optionally) storage.

The wired-network infrastructure includes a one-way feed of the campus network traffic, from a span port on one of the backbone routers, into a server located in the central computing facility. The research goal was to install and evaluate various network intrusion-detection systems, including several developed by Dartmouth researchers under previous projects, to evaluate their real-time performance on huge traffic flows.

Needless to say, the installation and operation of such an infrastructure requires careful planning and communication with the relevant campus departments. Although we had been collecting Wi-Fi network traces since 2001, the new wireless-network infrastructure was going to capture an order of magnitude more data, and the new wired-network infrastructure was going to capture data that had never been captured for research purposes at Dartmouth. Furthermore, the physical installation of over 200 Wi-Fi access points in about 10 large buildings around campus meant drilling holes into walls, leading to potential concerns about aesthetics.

The first step was to obtain permission from the Network Services group within Peter Kiewit Computing Services (PKCS), the central campus IT office. This step was easy, because we had developed the concept in collaboration with Network Services. We are fortunate to have a group of talented professionals who are also enthusiastic collaborators with researchers. We have repeatedly heard from colleagues, however, that this hurdle is very difficult in their organizations.

The next step was to obtain formal permission from Dartmouth's

Committee for Protection of Human Subjects (CPHS). CPHS services as the Institutional Review Board (IRB) for Dartmouth; all universities with federal research funding are required to operate an IRB so that research involving human subjects can be evaluated to ensure that risks are acceptable and subjects provide informed consent where appropriate. Our Wi-Fi network tracing effort was approved by CPHS several years prior, and our proposed effort was a subset of what we had done earlier, so a simple renewal was sufficient. Our wired-network effort was new, however, so we submitted a new project application to CPHS.

Meanwhile, we set out to obtain permission from the department heads located in the buildings where we hoped to place Wi-Fi sniffers. These buildings included the main library complex, the school of engineering, the school of business, a gymnasium, a student center, several dormitories, and several academic buildings. In some cases, we chose sites where renovation was underway and our sniffers (and their wiring) could be easily installed in the construction process, requiring less cost and no inconvenience to the building residents. In all cases, we met personally with the lead staff in each department, describing what we planned to do. We walked through their building, sometimes repeatedly, discussing in detail the placement of sniffers and their wiring. Each building required several months of planning to obtain permission, choose sites, confirm the sites with department staff, obtain quotes from electricians, install the wiring, and install the sniffers.

During this process, Sicker et~al. published a paper on the legal issues involved in network trace capture~\cite{sicker:legal}. The paper provides a thoughtful review of the many issues involved, and yet concludes that the legal status of network trace-collection for research purposes is not entirely clear. We consulted with the university counsel in depth, and with outside consultants, concluding that such trace collection could proceed as long as the research activity was closely coupled with network-operations activity. We had involved PKCS Network Services from the start, but we adjusted our research program to more directly meet their needs; our trace-capture facility can now support both operational and research goals simultaneously.

Furthermore, because of the scale of this effort, and the sensitive issues related to the privacy of network users, we met with several leadership groups on campus to explain our plans, answer their questions, obtain their feedback, and ultimately seek official approval from the College to proceed with trace collection. In particular, we met with the high-level faculty committee responsible for sponsored research and the provost-level council that includes all campus deans. In both cases we obtained valuable feedback that helped us to clarify our operating parameters. We developed an increasingly crisp understanding of the privacy risks and our mechanisms for mitigating those risks.

Ultimately, we decided to conduct a careful, objective study of our trace-collection infrastructure and our privacy-protection mechanisms. The College hired an outside expert, a researcher with several years of network-tracing experience in academic settings, to visit campus, interview the research teams, and to study our trace-collection infrastructure in detail. This visit served as a tremendous help to us, providing a critical eye to help us recognize where our plans could be improved or become more specific. In the end, based on the expert's advice and internal deliberations, the College leadership decided that the risks posed by the wired-network infrastructure (given the type of data needed by the researchers) were not easily mitigated, and the wired-network capture will not proceed. For the Wi-Fi infrastructure, we decided to add additional layers of security--- to ensure that the infrastructure itself can not be compromised by attackers--- and additional layers of encryption and inline anonymization to protect the privacy of network users. If, in the future, we make non-trivial changes to our data-collection infrastructure we will again ask the expert to evaluate our plans.

An important part of the process is communication and public notice, especially since informed consent is not feasible in an open wireless network covering numerous buildings and a shifting population. Every one of our sniffers is labeled with the URL of our website describing the project. We are posting notices at the entries to each building, informing visitors of the data-collection effort and directing them to the website for further information. At the request of the library, we are posting notices on every table in public areas of the library. Finally, we issued a press release describing our research and the scope of the data collection.

At this writing, our Wi-Fi sniffing infrastructure is nearly ready for operation. We include several layers of security on the sniffers, including extremely limited services, narrow firewall openings, no crypto keys in persistent storage, and frequent defensive port-scans. We discard all but the MAC layer from each frame, then encrypt each packet of captured frames before sending them to the server; at the server they are decrypted and immediately anonymized before being used for inline analysis or storage for offline analysis. The anonymization map is generated anew for each experiment, using a random seed that is discarded after use. As a result, very little sensitive information is captured and the most sensitive components (MAC addresses and SSIDs) are thoroughly anonymized.

The result is, we expect, a highly secure, privacy sensitive, scalable capture system for Wi-Fi networks, larger and more secure than any other ever assembled. We intend to use the infrastructure to collect operationally useful data for Network Services, and to serve our own ongoing research in trace anonymization techniques.

Reading list.

```
@INPROCEEDINGS{sicker:legal,
  author =      {Douglas~C. Sicker and Paul Ohm and Dirk Grunwald},
  title =       {Legal issues surrounding monitoring during network
research},
  booktitle =   procofthe # {ACM SIGCOMM Conference on Internet
Measurement (IMC)},
  year =        2007,
  pages =       {141--148},
  DOI =         {10.1145/1298306.1298307}
}

@INPROCEEDINGS{allman:etiquette,
  author =      {Mark Allman and Vern Paxson},
  title =       {Issues and etiquette concerning use of shared
measurement data},
  booktitle =   procofthe # {ACM SIGCOMM Conference on Internet
Measurement (IMC)},
  year =         2007,
  pages =        {135--140},
  DOI =          {10.1145/1298306.1298327}
}

@TechReport{burstein:culture,
  author =       "Burstein, Aaron~J.",
  title =        "Toward a Culture of Cybersecurity Research",
  institution =  "UC Berkeley Public Law Research Paper",
  year =         2008,
  number =       1113014,
  URL =          "http://ssrn.com/abstract=1113014",
}
```