;login:
DECEMBER 2014 VOL. 39, NO. 6
ELECTRONIC SUPPLEMENT

Conference Reports

FOCI '14: 4th USENIX Workshop on Free and Open Communications on the Internet

August 18, 2014, San Diego, CA

Summarized by Ben Jones and Eric Wustrow

Circumvention Technology

Summarized by Eric Wustrow (ewust@umich.edu)

ReClaim: A Privacy-Preserving Decentralized Social Network

Niels Zeilemaker and Johan Pouwelse, Delft University of Technology

ReClaim is a peer-to-peer replacement for online social networks. In ReClaim, friends who share friends can construct social graphs and send broadcast-encrypted messages to one another, but users who do not have overlapping friend circles cannot decrypt or send these messages.

Using a distributed hash table for friend lookups would reveal metadata of what friends you are interested in (i.e., what keys you look up in the DHT). Instead, ReClaim finds friends using private set intersection and homomorphic encryption, allowing friends to find each other in the peer-to-peer network without revealing their social graph to everyone.

ReClaim was evaluated using a Facebook data set as an existing social network, and found that it takes about 10 minutes for ReClaim to converge, finding about 95% of your friends through building up your friend graph through your existing friends' graphs.

Someone asked how you could implement "suggested friends" like Facebook currently implements under ReClaim. Niels Zeilemaker answered that it's not required to use private set intersection; you can allow other algorithms like the union or some combination to be used instead. In other words, a lower threshold of overlap could be used.

Another question was whether attribute-based encryption needs a centralized system or not. Zeilemaker answered that revocation of an attribute from a peer makes decentralizing ABE hard. The question was ultimately taken offline. Another person asked about how using a DHT exposes the social graph and whether it was possible to perform an anonymous lookup in a DHT. Zeilemaker answered that in a DHT, your data is stored by strangers instead of by your friends. However, the questioner pointed out that the data is encrypted, so it doesn't matter.

Another person asked whether it's possible to use existing social networks to help bootstrap this network. Zeilemaker answered that the existing social networks don't have the incentive to let people "free-ride" on top of their system.

TRIST: Circumventing Censorship with Transcoding-Resistant Image Steganography

Christopher Connolly, Patrick Lincoln, Ian Mason, and Vinod Yegneswaran, SRI International

Chris Connolly explained that TRIST attempts to subvert state-level censorship by encoding hidden messages steganographically in images and video. However, because images and video are posted to social networks (Facebook or YouTube), naively encoded hidden messages are often lost in the transcoding/compressing done by the social networks. TRIST attempts to solve this by using an encoding that survives this type of re-encoding. The encoding is even robust against resizing and filtering (e.g., sharpening filter). The bandwidth of this encoding is dependent on the high-frequency textures present in the cover image.

It's possible to extend this work to MPEG without much modification, and Connolly's presentation included a demo. It may also be possible to apply this to MPEG I-frames with some modification as well. TRIST is available on GitHub and integrated with StegoTorus.

Someone asked about comparing this to watermarking schemes and other related work, and whether watermark detection can be used to detect this as well. Another person asked whether the threat model of nonhuman adversary is fair. Another person asked whether there was a way to look at distribution of cover images vs. encoded images in order to distinguish encoded images automatically. Connolly answered that this type of analysis can be expensive to perform.

Facade: High-Throughput, Deniable Censorship Circumvention Using Web Search

Ben Jones, Sam Burnett, Nick Feamster, Sean Donovan, Sarthak Grover, Sathya Gunasekaran, and Karim Habak, Georgia Institute of Technology

Ben Jones stated that Facade aims to provide censorship circumvention over HTTP by encoding Web requests of censored users over HTTP search URLs (e.g., http://search.com/?q=correct+horse+battery+staple). Facade encodes your query into the search terms put into the HTTP GET request to the search engine. The search engine is also a Facade server and decodes the query string to get the true message (or request fragment) from the client. The server then reassembles the hidden messages given by the user (through search queries), requests the page the user requested, and encodes the result into an image served back to the user.

Someone wondered whether Facade is similar to Bananaphone, which encodes arbitrary data into English sentences. Another person asked about resource exhaustion (DoS), and Jones said that was not something they looked into. Someone else asked about some other kind of distribution magic that might be better than using a dictionary as the encoding function.

Theory and Policy

Summarized by Ben Jones (bjones99@gatech.edu)

Catching Bandits and Only Bandits: Privacy-Preserving Intersection Warrants for Lawful Surveillance

Aaron Segal, Bryan Ford, and Joan Feigenbaum, Yale University

Aaron Segal opened his presentation by describing the current state of the art in lawful surveillance; surveillance is implemented with secret processes, and the public is asked to trust the government without any evidence. Segal pointed out that this structure presumes an absolute tradeoff between national security and personal privacy. As an alternative, Segal goes on to show that surveillance with privacy protection is realizable with cryptographic guarantees and open processes for data collection. Unlike existing systems, his system achieves his privacy principles: openness, division of trust, enforced scope limiting, sealing time and notification, and accountability.

To demonstrate these problems, Segal presented a case study of the high country bandits. The high country bandits were a criminal group who were caught when the police dumped the data for 150,000 users from three cell towers and identified the single cell phone that was in all three locations. This use case did not satisfy any of Segal's principles since a single organization conducted the search and collected the data for 149,999 innocent users.

In response, Segal presented an existing private set intersection protocol adapted to this use case. This protocol achieves Segal's principles because it has openness (everyone knows the protocol's details), division of trust (no agency can do it alone), scope limiting (any agency can stop the protocol if too many users will be identified), a sealing time (all agencies get the final data), and accountability (no agency can execute without others). Using this protocol, Segal showed that it is possible to offer functionality similar to existing surveillance without sacrificing personal privacy.

Segal answered several questions about the work related to extensions for functionality such as dealing with situations where one group is not cleared for the results or offering more flexible matching on the intersection determination.

Symmetric Disclosure: A Fresh Look at k-Anonymity

EJ Infeld, Dartmouth College

EJ Infeld gave a fresh look at k-anonymity in terms of the links between groups. Infeld opened by defining her goal: to obscure the social graph of users. k-anonymity is a way to obscure the social graph where the size of the anonymity set is at least k. Infeld pointed out that we also need to care not only about the groups in the social graph, but the projection graph with links between groups. She highlighted that the problem here is that links between groups are extremely unlikely, so if two users are friends, then their communication patterns are clearly visible from the database.

Next, Infeld explored how we could increase the size of groups to offer better security for the social graph. She went on to note that the probability that two users will have cover traffic for their communication is roughly equal to the probability that there is a connection between two random groups. This means that if we want there to be a 20% probability that Alice and Bob have cover traffic, then we will need groups of 916 users.

Finally, Infeld proposed that if we would like to build a communication system in which groups would be big enough to guarantee cover traffic, we should explore ways of avoiding broadcasting a message to the entire group. She suggested that one such way would be to limit message retrieval to communications that come from a group in which a user has a friend, since this does not reveal new information about the social graph, as long as different requests from one user are not correlated.

An Internet with BRICS Characteristics: Data Sovereignty and the Balkanization of the Internet

Dana Polatin-Reuben and Joss Wright, University of Oxford

After the Snowden revelations, the role of the US and the Five Eyes nations in the Internet is being questioned. In response, BRICS nations (Brazil, Russia, India, China, South Africa) are considering data localization to prevent surveillance. If the BRICS nations were to take a strong approach to data sovereignty and localize traffic, then the fragmentation, or balkanization, of the Internet may be a real possibility. In the remainder of the talk, Dana Polatin-Reuben discussed the data sovereignty needs of each of the BRICS nations and whether each nation is leaning towards a single-stakeholder model with data localization, or a multi-stakeholder model where data is not localized.

Brazil offers a middle ground between Western countries and the BRICS nations with its strong push for privacy within the country contrasted with a commitment to a multi-stakeholder model at international conferences. In Russia, domestic surveillance has been in place since the mid-1980s, and the Russian government seems to be concerned with the control of information not just data. However, Russia seems to recognize the economic value of the EU's multi-stakeholder model. India's data sovereignty approach is focused on safeguarding its business process outsourcing sector, not on protecting its own citizens. This means that India is more likely to take a Western multi-stakeholder model.

China has seen democracy in Internet governance as equal representation between states, not representing the interests of individual citizens. China's approach has been focused on protecting its own trade secrets, but does not appear to be against the multi-stakeholder model due to its involvement with multi-stakeholder conferences and its collaboration with the US and EU. Finally, South Africa lags significantly behind other BRICS countries in setting an agenda and approved its cybersecurity framework as late as 2012. Therefore, it is unclear how South Africa will lean in the debate.

Polatin-Reuben concluded by noting the possible outcomes of the BRICS nations deciding on weak or strong data sovereignty. If BRICS nations select a weak approach, then there will be private-sector-led data protection, requiring China and Russia to give ground on cultural issues for economic incentives. If the nations adopt a strong approach, then a state-led approach would prevail, directly opposing a Western consensus, with strong economic implications.

Measurement and Analysis

Summarized by Ben Jones (bjones99@gatech.edu)

Global Network Interference Detection over the RIPE Atlas Network

Collin Anderson, University of Pennsylvania; Philipp Winter, Karlstad University; Roya, Independent Researcher

Anderson began the talk with a brief history of measurement, motivating the need for measurements from multiple vantage points on all the ISPs in a country. As a solution for some of these problems, Anderson et al. conducted measurements from the RIPE Atlas network, a home network measurement platform. RIPE Atlas is a good platform for censorship measurement because Atlas offers high diversity, push measurements, and the closest platform to interference at scale.

Anderson moved on to motivating several censorship measurement lessons with case studies from Turkey and Russia. In Turkey, Anderson showed the complexity of the censorship landscape and the nuances of censorship illuminated with larger data sets. In Russia, Anderson showed the importance of contextualizing results and digging deeper to understand what is actually going on.

Anderson concluded with several lessons and open questions about censorship measurement. He highlighted the importance of knowing the cultural context of censorship to understand what measurements actually mean. He also noted that censorship is often changed without warning, that measurement validation requires client environment documentation, and that data collection should be longitudinal. Finally, Anderson mentioned that the ethics of censorship are difficult to reason about because there is no framework for working through these issues.

Towards a Comprehensive Picture of the Great Firewall's DNS Censorship

Anonymous

Vern Paxson presented this talk in his role as PC co-chair on behalf of the authors of this paper, who wished to remain anonymous. Vern opened by noting that the Great Firewall of China (GFW) lacks comprehensive study. This paper addresses the problem by investigating the GFW's DNS injectors and trying to answer the following questions: What does the censor ban? How geographically pervasive is the coverage? Where is this topologically happening? Does this fail? Can we estimate the volume of censored traffic?

To measure the GFW, the authors made several measurements that should not return benign responses. The authors sent DNS requests that are not supposed to get responses to non-responsive addresses or by performing traceroute-like, TTL-limited, queries. In other cases, the authors sent requests for non-existent domains or conducted indirect scanning between open resolvers.

Using these techniques, the authors found several interesting results. In answer to the question of pervasiveness, the authors queried 150,000 open resolvers in China and only found 80 clean resolvers. All but one of these nodes appeared to use some technique to get the correct responses and one resolver appeared to be free of censorship. The authors determined that GFW nodes are deployed on the edge of China's network and also found that of the 130 million domains they tested, 35,000 were censored. Using binary search over portions of censored queries, the authors found that the censor was using suffix or substring matching and there were only 15,000 keywords. Because many uncensored domains have overlapping substrings/suffixes, this model leads to collateral censorship of other domains.

Finally, the authors estimated the volume of queries to one censorship node was around 2800 total requests per second with a clear diurnal pattern. However, this measurement is only from one node and there could be 100s to 1000s of such nodes in the country.

Counting Packets Sent Between Arbitrary Internet Hosts

Jeffrey Knockel and Jedidiah R. Crandall, University of New Mexico

In this talk, Knockel described a new side channel to count the number of packets sent between arbitrary hosts on the Internet and discussed potential mitigations of this attack. As background, IP fragmentation works by keeping fragments in a cache until all fragments arrive. Fragments are assigned to a cache on the basis of the IP ID field in the IP header. Linux assigns the IP ID field according to a counter that the kernel maintains independently for each destination address.

Knockel moved on by outlining a new side channel using IP fragmentation. To detect whether two hosts on the Internet are communicating, we first infer the value of the Linux machine's IP ID counter for target machine x. We begin by planting canary IP fragments on target machine x that consist of half of an echo request. Next, we spoof ICMP echo requests to target machine y spoofed from machine x. Later, we send the other half of the canary IP fragments to machine x and see whether the canary fragments are still there. If some of the canary fragments have been knocked out, then some of our canaries matched the value of the IP ID counter. By planting canaries across ranges of possible IP ID values, we can infer the value of the counter. Then we can see how many packets have been sent to x by measuring changes to the value of the counter.

To address attacks using this side channel, Knockel et al. contacted the Linux kernel team to patch the problem. The side channel is largely ameliorated in the current release and several old releases of the Linux kernel, but the problem fundamentally still exists. IP IDs from any RFC-compliant machine will always leak information about other packets sent because of the requirement that IDs be unique for every in-flight packet.

Security Audit of Safeplug "Tor in a Box"

Anne Edmundson, Anna Kornfeld Simpson, Joshua A. Kroll, and
Edward W. Felten, Princeton University

Anne Edmundson discussed Safeplug, a device that users plug into their home router to proxy their traffic through Tor. Because 86% of users have tried to become more anonymous online, Edmundson et al. explored how well Safeplug meets its claims, compared its security to alternatives like the Tor Browser Bundle (TBB), and determined whether there was any value in Torifying proxies.

Unfortunately, Edmundson et al. found numerous problems with the device in terms of usability and potential attacks. Disconcertingly, the terms of service for the device were extremely hard to find, a large problem for a device designed to provide security to those who know nothing about security. Additionally, the device's UI was helpful but contained a number of dangerous options for security-unaware users, like letting users become an exit node for Tor. Finally, Edmundson et al. also found a number of other security flaws in the device, like the same SSH root password for all versions of the device, using old versions of software, and failing to verify updates.

Edmundson concluded by noting some of the fundamental limitations of using a Torifying proxy in place of Tor Browser Bundle (TBB). Modern browsers leak lots of identifying information, so users lose a lot of privacy by using their own browser in place of a hardened one. However, Edmundson et al. concluded that while TBB is clearly superior, options like Safeplug can offer some additional privacy protection, although it is unclear how much.

Edmundson et al. answered several questions about Safeplug, but no consensus was reached. Specifically, the size of Safeplug's deployment is unclear and while they are clearly legally liable for putting people at risk with unreadable terms of service, the extent of this liability is unclear.