Research Statement

David Isaac Wolinsky
isaac.wolinsky@gmail.com http://www.davidwolinsky.com

My approach to research focuses on practicality and usability. In the Dissent project, I led the designing and construction of a scalable anonymous communication system with provable and quantifiable security properties, based firmly on theoretical research. As the technical lead in the Grid Appliance and IPOP projects, I created a distributed grid computing environment and a decentralized virtual private network, intended for and successfully used by lay persons. INDaaS (Independence-as-a-Service), a project under my advisement, detects and identifies potential failures in cloud and data center environments before they occur. My research efforts produce usable prototypes, that have been used for further research and collaboration. In the following sections, I elaborate on these research projects and conclude with future directions.

Dissent – Anonymous Group Communication

Over the past 30 years, the theoretical research community has made great strides in developing anonymous communication protocols offering strong anonymity. In contrast, the systems research community has made few inroads in deploying practical anonymity systems that offer *provable* and *quantifiable* security properties. At Yale, I joined the Dissent team with goals of producing a practical anonymity system with these guarantees [3, 13] using a theoretical model called Dining Cryptographers Networks [1] (DC-nets).

In my first effort, I designed and built the first scalable DC-net system [17, 18], supporting over 5,000 concurrent participants, two orders of magnitude more than previous approaches, in experiments that were ultimately limited by the testbed rather than the system's scaling potential. Unlike every earlier Dissent approach, we moved away from a fully connected model to a client-server model called anytrust that offers significantly better scalability with nearly the same strong security properties. A client in the anytrust model assumes there exists a single honest server but need not know which server to trust. This prototype also includes Neff's verifiable shuffle [12], which retains the anytrust notion during the bootstrapping of the scalable Dissent protocol.

Dissent uses a reactive mechanism to find disruptors, taking upwards of 60 minutes for a 1,000 member group. Traditional DC-nets and similarly Tor gain speed over mix-nets by making use of symmetric cryptography. In Verdict [4], I led the design and implementation of the first DC-net system to offer proactive accountability by constructing ciphertexts using asymmetric cryptography with provable correctness using zero knowledge proofs. Asymmetric cryptography, in contrast to symmetric cryptography, significantly impacts performance. As a middle ground, I merged these two approaches in a Dissent hybrid that normally utilizes symmetric cryptography and falls back to asymmetric cryptography during denial of service attacks.

My third major research contribution, Buddies [23], addresses a long-standing issue in all anonymous communication systems: the intersection attack [9]. The intersection attack correlates user activity to a set of anonymous, linkable messages in order to identify the owner. In Buddies, users mirror each other's activity, preventing the adversary from differentiating them. Other work recognized this as k-anonymity, yet only discussed its utility avoiding practical challenges, such as, obtaining and maintaining k-anonymity sets. Buddies' solves this challenge by simulating an intersection attack and applying user-defined policies to inform the anonymous communication system which users to treat as online or offline in order to maintain desired anonymity levels.

As technical lead in Dissent, I have also led and contributed to related projects on anonymous and biometric authentication. I have led students in building CryptoBook [11], which takes an existing federated login system, like Facebook, to create anonymous identities who remain anonymous even among users who have not used CryptoBook. DAGA, or deniable anonymous group authentication, combines anonymity, linkability, and deniability, so that users can authenticate across multiple sessions without fear that his compromised key could connect him back to any of his authentications. The biometric authentication project supports biometric authentication without ever divulging the user's biometrics to the authenticating source.

Dissent's software prototype has evolved from a simple evaluation tool into a product entering DARPA's technology transition, where it will likely be used in future defense technology projects. Over the past 4 years during my oversight, Dissent has been reviewed by two independent DARPA-funded teams, an analysis team and a red-team. Dissent has also been the cornerstone of several undergraduate senior design projects and graduate class projects both directly under my supervision and abroad.

IP over P2P – Restoring Any-to-Any Connectivity to the Internet

The Internet abounds with network asymmetries due to network address translation (NAT) and firewalls. Asymmetries inhibit direct sharing of content without using third-party services. IPOP [7] solves these asymmetries without dedicated third-parties through the use of a decentralized, structured overlay network, Brunet, allowing any user on any network to connect to any user on any network.

My first efforts [21] sought to make IPOP practical, to run without OS modification and minimal user configuration. The three core contributions were support for decentralized DHCP using a distributed hash table, a mechanism that enabled resources using a VPN and sharing a common LAN to communicate directly bypassing VPN overheads, and, finally, VPN support for virtual machine migration.

Decentralized systems require a dedicated bootstrapping platforms. Dependence on this dedicated platform inhibits personal use, as a result users often join existing infrastructures. Internet asymmetries, such as NAT, only further exasperate the problem. To eliminate this dependency, I recognized [22] that a decentralized bootstrapping system required reflection for peers to discover their own network mappings, rendezvous for peers to find other peers, and relaying to exchange network mappings. I built a prototype system supporting these features that took advantage of existing public overlays including XMPP (Google Talk), BitTorrent's DHT (Kademlia), and even public instances of IPOP.

Securing IPOP remained as the precluding issue to adoption. Users frequently browse the web insecurely, as witnessed by the many users who accept invalid certificates. On the other hand, users value and understand social networking websites. I exploited this in building GroupVPN [20] and SocialVPN [8, 5], enabling users to establish trusted, secure connections without ever explicitly exchanging certificates.

IPOP has had substantial impact, primarily in the form of SocialVPN [8] and GroupVPN [20], both of which, during their heydays, had hundreds of active users running independently from our infrastructure. These individuals included other researchers, companies, and even individuals playing video games. IPOP has also played a fundamental role in Grid Appliance as the networking substrate.

Grid Appliance – Decentralized Grid Computing

Research progress has often been limited due to lack of computing resources. Many "@home" projects, e.g., curing cancer or finding extraterrestrial life, recruit volunteers for their computing resources, however, limiting these resources for project specific purposes. Fortunately, most researchers have some computing infrastructure, which experience bursty use as a paper submission looms. I built Grid Appliance [15, 19, 16] with this intuition allowing researchers to volunteer their idle computing cycles with each other.

Connecting resources under different administrative domains resulted in many challenges for Grid Appliance: establishing networking between all systems, safe sharing of resources, and ensuring fairness. Networking issues led to the design and development of IPOP and GroupVPN, which Grid Appliance uses as a means to federate access to resources. Users within a common GroupVPN have access to all resources, which further divides into subgroups or institutions that offers privileged access to resources for others within that subgroup. The Grid Appliance sandbox uses virtual machines and networking, significantly limiting Grid Appliance's attack surface. Virtualization also provides a homogeneous environment, so that researchers need to build only a single Grid Appliance compatible binary. I guided efforts to extend Grid Appliance to support on-demand bootstrapping both Hadoop and MPI clusters.

Grid Appliance requires a dedicated resource for managing the pooled resources. I collaborated on efforts [10] that removed this restriction, revealing a completely decentralized solution to constructing compute cluster pools on demand (pond) using a decentralized distribute and aggregate platform.

Grid Appliance experienced much success. In 2008, the National Science Foundation funded Archer, a computer architecture research community built around Grid Appliance [6]. At its height, Archer spanned over a dozen university and nearly a thousand computers. Many other universities adopted Grid Appliance as an educational platform for distributed computing, and several external organizations used it as a distributed computing solution. I made significant contributions to Grid Appliance: designing the entire framework, maintaining it thereafter, administering Archer, and leading many tutorials at universities and eonferences.

Independence-as-a-Service – Improving Cloud Reliability

Using the cloud has become a popular business clichè, as companies increasingly move their computation and data away from privately owned resources into the cloud. Caution abounds, rightly so, as clouds obscure their internals leading to reliability concerns. Many examples in recent history have shown that reliance on a

single cloud can still result in loss of service. For example, on Christmas Eve 2012, Netflix customers in North America lost access, while those in Europe remained unaffected [2]. Netflix builds their infrastructure using Amazon; however, despite their best efforts, Amazon cannot perfectly isolate each of their availability zones. Netflix may have been able to avoid this unfortunate situation had it been aware of hidden dependencies within Amazon or chosen additional services providers.

The umbrella project that I oversee, called Independence-as-a-Service (INDaaS) [25], performs a three step procedure for assisting application providers in optimizing their cloud deployments for reliability as well as informing cloud providers of their own blemishes. In step 1, the INDaaS performs resource discovery, identifying all the hardware and software components within a cloud. In step 2, the INDaaS builds relationships or dependency graphs akin to those in fault tree analysis [14]. Finally, the INDaaS performs analysis across similar deployment plans that use one or more clouds in order to recommend a highly available deployment plan. To address privacy concerns during step 3 when performed across multiple cloud providers, iRec [24] uses private set intersection cardinality, offering privacy preservation but allowing recommendations that span multiple clouds.

Future Work

Short-term Work on Dissent brought my attention to the general problem of constructing decentralized, secure group systems. There exist many protocols and systems to address this, but in order to be made practical, they make security compromises in their construction, such as assuming an honest leader or a single server. An anonymous communication system with such an assumption is worthless. I am currently designing a generalized framework to deal with this. The major problems identified thus far include constructing a group definition, system randomness, and handling client rejoins.

Short-term Buddies offers a mechanism for maintaining anonymity by monitoring pseudonyms, such as, short-lived web browsing sessions and long-lived Twitter usernames. The purely decreasing nature of anonymity hampers the long-term use of a pseudonym. To address this, I am investigating the use of anonymous reputation to enable users to build up reputation scores, as an alternate to a pseudonym, with the caveat that reputation score not be linkable over time and not subject to decreasing anonymity.

Medium-term My cloud research brought my attention to a general problem experienced in all my development projects: debugging and understanding distributed systems. Most cloud failures occur because hardware failures trigger unexpected behavior in software. I plan to build a system to address these issues that automates the testing of distributed systems including cloud platforms. The system will capture both system and network state for each outgoing and incoming message using network and system virtualization. When the distributed system encounters a bug, these stored states can be used as a trace through the system revealing the source of the bug.

Long-term Google's Eric Schmidt stated, "We can end government censorship in a decade. The solution to government surveillance is to encrypt everything." I plan to accomplish this goal, but believe that encryption to be insufficient. Encryption does not solve the meta-data problem, the who's, what's, and where's, rather it hides the details, Secondly, it does not address network-level censorship. I believe the best approach to eliminating this threat is through encryption and anonymity, through a project I am developing called Secure, Anonymous Internet (SAI). An adversary hoping to censor SAI has only one choice: complete shutdown of the Internet.

As we introduce more complex systems into our lives, tensions between usability and other system properties will continue. I am a systems researcher, but I enjoy diversity, I have had a wide range of experiences including distributed systems, networking, security, and privacy. My research has been funded by DARPA, NSF, and industry grants, sources that will have funding opportunities for this area of research in the future. Beyond continuing my own research efforts, I am excited by the opportunity to work with a diverse faculty in order to find new, important, and interesting problems that can be solved in the scope of systems research.

References

[1] D. Chaum. The Dining Cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, pages 65–75, Jan. 1988.

[2] A. Cockcroft. `http://techblog.netflix.com/2012/12/a-closer-look-at-christmas-eve-outage.html`, 12 2012.

[3] H. Corrigan-Gibbs and B. Ford. Dissent: accountable anonymous group messaging. In *17th ACM conference on Computer and communications security (CCS)*, Oct. 2010.

[4] H. Corrigan-Gibbs, D. I. Wolinsky, and B. Ford. Proactively accountable anonymous messaging in Verdict. In *22nd USENIX Security Symposium (USENIX Security)*, Aug. 2013.

[5] R. Figueiredo, P. O. B. Boykin, P. St. Juste, and D. Wolinsky. Social VPNs: Integrating overlay and social networks for seamless p2p networking. In *Workshp on Collaborative Peer-to-Peer Systems (COPS)*, 2008.

[6] R. J. Figueiredo, P. O. Boykin, J. A. B. Fortes, T. Li, J. Peir, D. Wolinsky, L. K. John, D. R. Kaeli, D. J. Lilja, S. A. McKee, G. Memik, A. Roy, and G. S. Tyson. Archer: A community distributed computing infrastructure for computer architecture research and education. In *CollaborateCom*, November 2008.

[7] A. Ganguly, A. Agrawal, O. P. Boykin, and R. Figueiredo. IP over P2P: Enabling self-configuring virtual IP networks for grid computing. In *International Parallel and Distributed Processing Symposium*, 2006.

[8] P. S. Juste, D. Wolinsky, P. O. Boykin, M. Covington, and R. Figueiredo. Socialvpn: Enabling wide-area collaboration with integrated social and overlay networks. In *Journal of Computer Networks*, 2010.

[9] D. Kedogan, D. Agrawal, and S. Penz. Limits of anonymity in open environments. In *5th International Workshop on Information Hiding*, Oct. 2002.

[10] K. Lee, D. Wolinsky, and R. J. Figueiredo. Pond: Dynamic creation of htc pool on demand using a decentralized resource discovery system. In *International Symposium on High Performance Distributed Computing (ACM HPDC)*, 2012.

[11] J. Maheswaran, D. I. Wolinsky, , and B. Ford. Crypto-Book: An architecture for privacy preserving online identities. In *Hot Topics in Networks*, Nov. 2013.

[12] C. A. Neff. A verifiable secret shuffle and its application to e-voting. In *ACM Conference on Computer and Communications Security*, pages 116–125, Nov. 2001.

[13] E. Syta, H. Corrigan-Gibbs, S.-C. Weng, D. I. Wolinsky, B. Ford, and A. Johnson. Security analysis of accountable anonymity in dissent. *ACM Transactions on Information and System Security (TISSEC)*, aug 2014.

[14] W. Vesely, F. Goldberg, N. Roberts, and D. Haasl. *Fault Tree Handbook*. U.S. Nuclear Regulatory Commission, Jan. 1981.

[15] D. I. Wolinksy, A. Agrawal, P. O. Boykin, J. Davis, A. Ganguly, V. Paramygin, P. Sheng, and R. Figueiredo. On the design of virtual machine sandboxes for distributed computing in wide area overlays of virtual workstations. In *VTDC*, 2006.

[16] D. Wolinsky, P. Chuchaisri, K. Lee, and R. Figueiredo. Experiences with self-organizing, decentralized grids using the grid appliance. In *Cluster Computing*, 2013.

[17] D. I. Wolinsky, H. Corrigan-Gibbs, B. Ford, and A. Johnson. Scalable anonymous group communication in the anytrust model. In *European Workshop on System Security (EuroSec)*, Apr. 2012.

[18] D. I. Wolinsky, H. Corrigan-Gibbs, A. Johnson, and B. Ford. Dissent in numbers: Making strong anonymity scale. In *10th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, Oct. 2012.

[19] D. I. Wolinsky and R. Figueiredo. Experiences with self-organizing decentralized grids using the grid appliance. In *International Symposium on High Performance Distributed Computing (ACM HPDC)*, 2011.

[20] D. I. Wolinsky, K. Lee, P. O. Boykin, and R. Figueiredo. On the design of autonomic, decentralized vpns. In *the 6th International Conference on Collaborative Computing (CollaborateCom)*, 2010.

[21] D. I. Wolinsky, Y. Liu, P. S. Juste, G. Venkatasubramanian, and R. Figueiredo. On the design of scalable, self-configuring virtual networks. In *IEEE/ACM Supercomputing 2009*, November 2009.

[22] D. I. Wolinsky, P. St. Juste, P. O. Boykin, and R. Figueiredo. Addressing the P2P bootstrap problem for small overlay networks. In *10th IEEE International Conference on Peer-to-Peer Computing*, 2010.

[23] D. I. Wolinsky, E. Syta, and B. Ford. Hang with your Buddies to resist intersection attacks. In *Conference on Computer and communications security (CCS)*, Nov. 2013.

[24] E. Zhai, R. Chen, D. I. Wolinsky, and B. Ford. An untold story of redundant clouds: Making your service deployment truly reliable. In *9th Workshop on Hot Topics in Dependable Systems (HotDep)*, Nov. 2013.

[25] E. Zhai, R. Chen, D. I. Wolinsky, and B. Ford. Heading off correlated failures through independence-as-a-service. In *10th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2014.