

Research Statement

Jeffrey Pang

My research interests span the broad areas of networking, distributed systems, and security. In particular, I enjoy devising techniques to improve pervasive computing environments and building Internet-scale distributed systems. All my research efforts are unified by an approach that combines empirically driven design, building usable prototypes, and collaboration with researchers in both industry and academia. I believe in this approach because it develops research that is both academically interesting and practical; thus far, my work is recognized by publications in top-tier networking and mobile computing conferences (several of which are already included in the curricula of networking and systems courses¹) and by industry interest in its application. I want to continue applying this methodology to new problems in the space of pervasive computing and distributed systems, but before I present my future research plans, I describe my current research efforts and their contributions.

1 Thesis Research

The CRA challenged the research community to “give computer end-users security they can understand and privacy they can control,” and it has responded with many privacy preserving applications. However, there is little prior work that addresses the privacy risks associated with the low-level network protocols that pervasive computing applications are built upon, and state-of-the-art wireless protocols, such as Wi-Fi and Bluetooth, only prevent eavesdroppers from obtaining the contents of message payloads. This leaves protocol control information exposed, which enables user tracking, profiling, and traffic analysis attacks that are outside the scope of end-user and application controls. In collaboration with my adviser, Prof. Srinivasan Seshan, and with researchers at Intel Research, the University of Washington, and the University of Colorado, my thesis work improves our understanding of the privacy of wireless protocols and demonstrates novel ways to build more private protocols. I highlight the major contributions of this research below.

Quantifying privacy threats. It is known that device addresses exposed in wireless protocols can be used to identify and track users. However, initial proposals to address this problem assume that changing these addresses periodically is sufficient to prevent tracking. Contrary to common knowledge, my work demonstrates that this assumption is wrong because many *implicit identifiers* often remain [Mobicom 2007, HotOS 2007]. Using practical machine learning techniques and analysis on the 802.11 traffic of hundreds of users, my work shows that an adversary can accurately track many users even if devices periodically change addresses and use state-of-the-art link layer encryption, such as WPA.

Identifier-concealing link layer protocols. The analysis above makes it clear that the most comprehensive way to thwart link layer attacks is to hide *all* explicit fields that can link *any* two packets together. This is challenging because many identifying fields play key roles in all wireless protocols today, such as device discovery and message filtering. My contribution is insight that enables complete and efficient wireless protocols that do not expose any explicit identifiers [MobiSys 2008, HotNets 2007]. For instance, a fundamental shift in protocol design is to perform authentication before discovery. To demonstrate the practicality of these ideas, I have implemented SlyFi, a new wireless protocol that maintains the features and efficiency of 802.11/WPA by using cryptographic primitives available on commodity hardware, but is also identifier-concealing. This work won the Best Paper Award at MobiSys 2008 and Intel is leading a push for its application and IEEE standardization.

Trustworthy service discovery. I am also examining novel ways to bootstrap trust between devices that have never communicated before. My first contribution in this area is Wifi-Reports, a service that enables users to evaluate Wi-Fi hotspots before they use them. Wifi-Reports is a novel reputation system for

¹ For example:

Cornell’s *Advanced Computer Networking* (CS 619, Fall 2004)
Princeton’s *Systems and Networking for Virtual Worlds* (COS-597B, Fall 2008)
Stanford’s *Networked Systems for Virtual Worlds* (CS340V, Fall 2008)
UIUC’s *Advanced Topics in Distributed Systems* (CS 525, Spring 2007-2008)
UT Austin’s *Wireless Networking* (CS386W, Fall 2008)
Wisconsin’s *Advanced Computer Networks* (CS 740, Spring and Fall 2008)

wireless networks that uses measurement reports contributed by users. Wifi-Reports mitigates the impact of fraud, and it is unique in that it preserves the location privacy of users that contribute, a crucial property of pervasive systems. This is achieved with a new cryptographic protocol that ensures reports are anonymous and unlinkable and limits each user to one report per network. My second contribution in this area is a set of privacy preserving mechanisms for personal devices (e.g., iPods and portal game stations) to discover each other [HotNets 2007]. These mechanisms leverage identity-based encryption and private matching techniques to discover unknown but transitively trusted devices — i.e., those in trusted domains or belonging to “friends-of-friends.”

2 Other Research

I also have a significant interest in large scale distributed systems. I have built and measured systems for massively multiplayer games, distributed file systems, and the Domain Name System (DNS).

Large-scale P2P games. In collaboration with Microsoft Research, I have designed and implemented major components of Colyseus [NSDI 2006] and Donnybrook [SIGCOMM 2008, IPTPS 2007]. These are peer-to-peer architectures that enable fast-paced multiplayer games with hundreds of simultaneous players, compared with the limit of 16 to 32 using architectures today. These architectures are the first peer-to-peer game systems that are validated using major commercial games (Quake II and III) and the first to demonstrate user satisfaction with real players and empirically derived workloads. The crucial challenge in these systems is to quickly and scalably deliver updates about objects to other peers that are interested in them (e.g., a player has to send updates about himself to those that see him). I designed several novel mechanisms for update delivery, including area-of-interest-based prefetching strategies, attention-oriented prioritization, and an overlay multicast scheme that can meet tight delay bounds and significant membership churn.

Distributed file systems. In collaboration with Intel Research, I have designed and implemented D2 [ICDCS 2007], a novel distributed file system that ensures files close together in the file system namespace are likely to be hosted on the same nodes, thereby improving availability and performance of user and application tasks over distributed hash table (DHT) based systems. This is because typical user and application tasks operate over many files so clients of DHT systems, which generally place files on random nodes, are dependent on many more nodes for their data than those of D2. An unexpected finding is that D2 can also balance load under real file system workloads without significantly more overhead than DHT-based systems.

Internet measurement. Finally, I performed measurement studies of DNS to better understand federated Internet infrastructure [IMC 2004a, IMC 2004b]. These studies involve active and passive measurements of over 300,000 DNS servers for several weeks. I have also analyzed measurements from a large study of BGP route selection, with a focus on inferring the causes of failures and inefficient routes [SIGCOMM 2004]. These studies are the first to discover several important properties about the Internet’s infrastructure. For instance, many DNS servers do not obey TTLs.

3 Future Work

I am interested in building protocols and systems that support ubiquitous computing applications in unmanaged and untrusted environments, such as home, travel, and social settings. These environments are increasingly littered with personal network devices and appliances such as laptops, music players, game stations, and set-top boxes. Thus, they will soon have many of the same management problems that enterprises have today. Unfortunately, existing solutions do not apply because these environments have different trust assumptions, have little or no centralized control, and lack professional IT staff. For example, even the most recent research proposals for diagnosing faults in enterprise Wi-Fi networks generally assume that a centralized IT department exists and that it can collect potentially sensitive data from all client devices. I believe my thesis work gives me the expertise to begin a new research agenda that addresses fault diagnosis, access control, measurement, and other critical aspects of network management in less trustworthy environments. I outline two ideas that may be part of this agenda below.

Private fault diagnosis and repair. The environments described above are typically externally managed. For example, users already use external vendors to troubleshoot personal computers (e.g., Best Buy’s Geek-

squad). However, using existing mechanisms designed for the enterprise, even professionals cannot diagnose many network problems without unfettered access to all devices therein. However, such access risks the privacy of users' data. For example, there are also reports of Geeksquad employees using their access to snoop on users' files. Troubleshooting networks currently requires even more invasive access to more devices and, often, continuous monitoring. Thus, there is an opportunity to devise new network, operating system, and device primitives that can cordon off private user information while still leaving enough exposed to accurately diagnose and address problems. This is challenging because, as I discovered in my thesis research, network protocols often unintentionally release sensitive information during their normal operation. One promising idea is to use taint-tracking techniques to trace and prevent unintended disclosures.

User and application-centric access control. Many applications of wireless devices operate over multiple mobile devices, possibly belonging to different users (e.g., games, media sharing, and peripheral extension). Managing cross-device access control today is typically *device-centric*. That is, user accounts and permissions on one device can not be generally be applied on another; devices declare trust in other *devices* (e.g., using Bluetooth pairing), rather than other *users* and *applications*. However, people typically think about who can have access to their files and resources in terms users and applications, not particular devices. I believe there is an opportunity for designing new user- and application-centric access control paradigms that unify access control across devices. The increasing acceptance of online social networks to mediate access to personal data demonstrates one promising avenue for managing access control on personal devices. One important research challenge is how to leverage transitive trust without unintentionally exposing social relationships, particularly when devices are disconnected. The insights that enabled me to use transitive relationships in private discovery protocols might also give rise to new ideas to meet this challenge.

References

- [HotNets 2007] **J. Pang**, B. Greenstein, D. McCoy, S. Seshan, and D. Wetherall. Tryst: The case for confidential service discovery. In *HotNets VI: The Sixth Workshop on Hot Topics in Networks*, Nov. 2007.
- [HotOS 2007] B. Greenstein, R. Gummadi, **J. Pang**, M. Y. Chen, T. Kohno, S. Seshan, and D. Wetherall. Can ferris bueller still have his day off? protecting privacy in an era of wireless devices. In *HotOS XI: Proceedings of the 11th Workshop on Hot Topics in Operating Systems*, May 2007.
- [ICDCS 2007] **J. Pang**, P. B. Gibbons, M. Kaminsky, S. Seshan, and H. Yu. Defragmenting dht-based distributed file systems. In *ICDCS '07: Proceedings of the 27th International Conference on Distributed Computing Systems*, June 2007.
- [IMC 2004a] **J. Pang**, A. Akella, B. Maggs, S. Seshan, and A. Shaikh. On the responsiveness of dns-based network control. In *IMC '04: Proceedings of the 2004 Internet Measurement Conference*, Oct. 2004.
- [IMC 2004b] **J. Pang**, J. Hendricks, A. Akella, B. Maggs, , R. D. Prisco, and S. Seshan. Availability, usage, and deployment characteristics of the domain name system. In *IMC '04: Proceedings of the 2004 Internet Measurement Conference*, Oct. 2004.
- [IPTPS 2007] **J. Pang**, F. Uyeda, and J. R. Lorch. Scaling peer-to-peer games in low-bandwidth environments. In *IPTPS '07: Proceedings of the 6th International Workshop on Peer-to-Peer Systems*, Feb. 2007.
- [Mobicom 2007] **J. Pang**, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall. 802.11 user fingerprinting. In *Mobicom '07: Proceedings of the 13th Annual International Conference on Mobile Computing and Networking*, Sept. 2007.
- [MobiSys 2008] B. Greenstein, D. McCoy, **J. Pang**, T. Kohno, S. Seshan, and D. Wetherall. Improving wireless privacy with an identifier-free link layer protocol. In *MobiSys '08: 6th International Conference on Mobile Systems, Applications, and Services*, June 2008.
- [NSDI 2006] A. Bhambe, **J. Pang**, and S. Seshan. Colyseus: A distributed architecture for interactive multiplayer games. In *NSDI '06: Proceedings of the 3rd Symposium on Network Design and Implementation*, May 2006.
- [SIGCOMM 2004] A. Akella, **J. Pang**, B. Maggs, S. Seshan, and A. Shaikh. A comparison of overlay routing and multihoming route control. In *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, Aug. 2004.
- [SIGCOMM 2008] A. Bhambe, J. Douceur, J. R. Lorch, T. Moscibroda, **J. Pang**, S. Seshan, and X. Zhuang. Donnybrook: Enabling large-scale, high-speed, peer-to-peer games. In *SIGCOMM '08: Proceedings of the 2008 conference on Applications, technologies, architectures, and protocols for computer communications*, Aug. 2008.