IEEE ComSoc Distinguished Lecture Tour – West USA, Dec. 3-14, 2014
Ying-Dar Lin, IEEE Fellow
National Chiao Tung University, TAIWAN
December 2014

DLT Planning and Itinerary

This is a DLT (Distinguished Lecture Tour) packed with six talks and piggybacked to IEEE Globecom in Austin, Texas. Among the six talks covering two topics, traffic forensics and software defined networking (SDN), three are to the academia arranged by myself and the other three are to the industry arranged by the IEEE Austin Chapter. My first talk was on traffic forensics at Naval Postgraduate School in Monterey, California, hosted by Prof. Preetha Thulasiraman who I have known for years. Then I drove to San Jose State University for my second talk, also on traffic forensics, hosted by Chair Xiao Su and Prof. Chao-Li Tarng who was my colleague when I had my sabbatical in Cisco, San Jose. I flew to Oklahoma City to have my third talk, on SDN, in University of Oklahoma, hosted by Prof. Krishnaiya Thulasiraman who has been my academic coach. Then I flew to Austin to attend Globecom and several editorial boards, program committees, and technical committees. After Globecom, the Chair of Austin Chapter, Dr Fawzi Behmann, accompanied me to three talks, 10AM, 2:30PM, and 6PM on Friday (12/12), at IBM Research and AT&T Labs, both on traffic forensics but couldn't be merged due to potential competition, and IEEE Austin Chapter, on SDN. At IBM and AT&T, the hosts were John Carter and Robert Dailey/Chris Chase, respectively.

I provided five topics (research roadmap driven by NBL, traffic forensics, benchmarking smartphones, open source for networking, SDN) for my hosts to choose from. A 4-min video is available at https://www.youtube.com/watch?v=BuxQ9Yk3OXc&feature=youtu.be. Thanks should go to the chairs of three local chapters who helped the local publicity. They are Santa Clara Chair, Rajeev Krishnamoorthy, Tulsa Chair, Pramode Verma, and Austin Chair, Fawzi Behmann, especially Fawzi for his excellent arrangement to the industry in Austin.

Six Lectures

The lecture on traffic forensics is on a series of 20 research papers of mine done in the past 5 years, ranging from traffic capture on the campus beta site, replay from captured traces, classification leveraging commercial devices, to detection and analysis of intrusions and malware with techniques of signature matching, statistical behavior analysis, and hybrid approach. The lecture on SDN is a tutorial and survey. I argued why, where, and when for SDN. Then I illustrated how SDN works in sections of standardization, development, and testing. For the first three lectures to the academia, it was the final week in universities in US. The attendees were mostly faculty members, post-doc researchers, and Ph.D. students. The number of attendees per lecture was about 20-25. For the next three lectures to the industry, the number

of attendees was 20, 20, and 40 in IBM, AT&T, and Austin Chapter, respectively. The number of questions asked was 5 and 20 per lecture, in the academia and industry. It appears that the industrial people were more curious on how various tricks were actually done. In summary, I received the positive comments like “very impressive work”, “very informative talk”, “very vivid arguments”, “it clarifies my doubts about SDN”, etc. Chris Chase, an AT&T Fellow, commented “this is the most interesting talk I had attended recently” and “we should call back our colleagues on vacation to listen to this”.

In-Depth Discussions During and After Lectures

The lectures triggered many questions from the audience. I list major questions and my answers below. For SDN questions similar to the ones listed in my previous DLT report, please refer to that report.

1. [Traffic Forensics] How automatic is traffic classification and extraction from raw traffic to classified library? How much time is needed to re-construct the PCAP Lib?

Most of the process has been automated, including traffic replay, identifying anchor packets from logs, associating and extracting sessions and flows, and anonymizing packets. But human inspection is needed on the results, especially the anonymized traces to make sure no leakage of privacy. On the average, two weeks would be needed to re-create a PCAP Lib.

2. [Traffic Forensics] What is the relationship between TTF (Time to Fail) and stability score (traffic volume divided by number of defects)?

They are positively correlated, but require further research from the aspect of software engineering.

3. [Traffic Forensics] Since defects caught by real traffic only tell us the phenomena, what are the root causes of those defects and how can we find them?

It is the vendor to find the root causes of defects with the phenomena of hanging, rebooting, connection failure, function failure, slow down, etc. We provide the right segment of traffic to help them reproduce and diagnose the defects.

4. [SDN] As SDN turns network appliances to apps, what apps will be created in the long run?

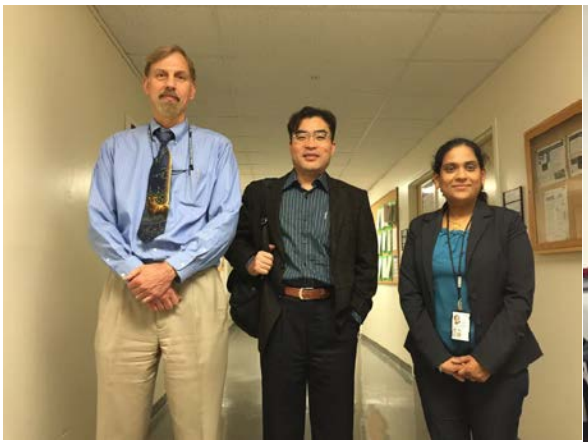
Each type of existing network functions (router, switch, Wi-Fi, VoIP, network security, 4G, etc.) will correspond to a series of apps, which might sum up to 50 apps. While the network functions are not limited to the form factor of an appliance, there might be 50 more apps created to provide value-added services and composite suites of services.

5. [SDN] Who are the existing providers of SDN switches, controllers, and apps?

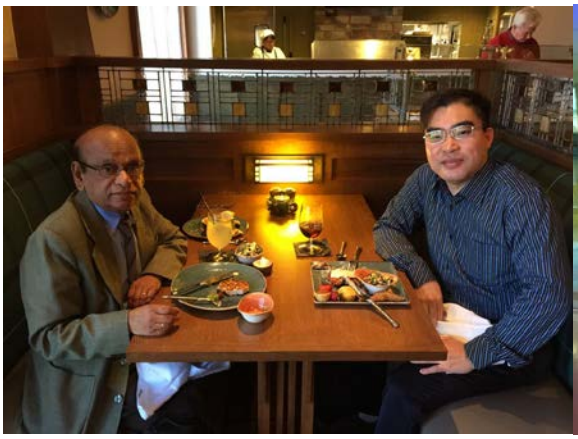
OpenFlow switch startups came first. Next, all traditional vendors are starting to roll out their switches. Most major controllers are open source projects, with some commercial module support. Some vendors, such as HP, Ericsson, and Cisco, claim to provide app store frameworks.

6. [SDN] What if the controllers running on virtual machines are too slow, compared to switches?

Switches respond in microsecond to nanosecond as required in the data plane, while controllers respond in millisecond to sub-seconds at the control plane. To speedup controllers, clusters of virtual machines or accelerated platforms (such as Intel DPDK and Open Data Plane - ODP) could be used.




Left: Naval Postgraduate School – Douglas Fouts, me, Preetha Thulasiraman
Right: San Jose State Univ. – Chao-Li Tarng, me, Xiao Su




Left: Univ. of Oklahoma - Krishnaiya Thulasiraman and me
Right: IBM Research Austin – Fawzi Behmann, me, John Carter, and two attendees


Left: AT&T Labs – me, Robert Dailey, Chris Chase, Fawzi Behmann
Right: Austin Chapter – attendees on Friday night

Appendix:

Talk Title: **Software Defined Networking: Why, When, Where, and How**

Abstract:

The first wave of cloud computing was to centralize and virtualize servers into the clouds, with a phenomenal result. The emerging second wave, named Software Defined Networking (SDN), is to centralize and virtualize networking, especially its control, into the clouds. SDN deployment started from data centers and now expands to the model of "networking as a service" (NaaS) offered by the operators to enterprise and residential subscribers. By centralizing the control-plane software of routers and switches to the controller, and its applications, and controlling the data-plane of these devices remotely, SDN reduces the capital expenditure (CAPEX) and operational expenditure (OPEX) because the devices become simpler and hence cheaper and number of administrators could be reduced. SDN also enables fast service orchestration because the data plane is highly programmable from the remote control plane at controllers and applications. However, as we detach control plane from where data plane resides, new protocols shall be introduced between control plane and data plane, as the southbound API between controllers and devices and the northbound API between controllers and applications. As we further extend the control plane from controllers to applications such as Service Chaining (SC) and data plane from devices to Network Function Virtualization (NFV), newer mechanisms and APIs need to be added to these APIs. We argue why, when, and where SDN would prevail, and then illustrate how to make it happen. We shall introduce the key technology

components, including OpenFlow, SC, NFV, and Network Service Header (NSH) and then review the issues on standardization, development, deployment, and research. At the end, the development and deployment experiences of a campus SDN solution for Wi-Fi/switch control and management are shared.

Talk Title: **Traffic Forensics: Capture, Replay, Classification, Detection, and Analysis**

Abstract:

If computer forensics is to identify, preserve, recover, and analyze who did what on a computer, network forensics is to do the same on a network. Compared to network forensics, which has wider forensics targets on devices (e.g., switches, routers, access points, firewalls, gateways) and packets between them, traffic forensics focuses on packets alone. When these devices are black boxes and do not have storage to record what happened, which are often true, traffic forensics then approximates network forensics. In this talk, we present a series of technologies and tools we developed to capture, replay, classify, detect, and analyze traffic. From the architectures of a beta site embedded into an operational campus network with live traffic, to replay captured traffic with stateless or stateful replayers in wired or wireless environments, we build the basic infrastructure and tools to play with real traffic. A case study is reported to see how effective the accumulated packet traces are in triggering bugs in products under development.

Then we present another class of techniques leveraging the domain knowledge of existing products to classify traffic into various applications or malicious intrusions and malware. A classified PCAP library, associated techniques, and their evaluation are illustrated. With these integrated, a case study is reported to redefine security criteria with functionality, robustness, performance, and stability testing, in order to complement existing criteria such as Common Criteria, ICSA, and NSS. As sources of intrusions are often malware carried in application payloads, collect, analyze, and detect malware are the essential ways to build the defense lines. Thus, we present the mechanisms to collect and analyze active and passive malware through honeypot and P2P, respectively. At the end, we present detection mechanisms for traditional malware, Android malware, and Advanced Persistent Threat (APT).

Autobiography:

YING-DAR LIN is a Distinguished Professor of Computer Science at National Chiao Tung University (NCTU) in Taiwan. He received his Ph.D. in Computer Science from UCLA in 1993. He served as the CEO of Telecom Technology Center during 2010-2011 and a visiting scholar at Cisco Systems in San Jose during 2007–2008. Since 2002, he has been the founder and director of Network Benchmarking Lab (NBL, www.nbl.org.tw), which reviews network products with real traffic. NBL recently became an approved test lab of the Open Networking Foundation (ONF). He also cofounded L7 Networks Inc. in 2002, which was later acquired by

D-Link Corp. His research interests include design, analysis, implementation, and benchmarking of network protocols and algorithms, quality of services, network security, deep packet inspection, wireless communications, embedded hardware/software co-design, and recently software defined networking. His work on “multi-hop cellular” was the first along this line, and has been cited over 650 times and standardized into IEEE 802.11s, IEEE 802.15.5, WiMAX IEEE 802.16j, and 3GPP LTE-Advanced. He is an IEEE Fellow (class of 2013), an IEEE Distinguished Lecturer (2014&2015), and a Research Associate of ONF. He is currently on the Editorial Boards of *IEEE Transactions on Computers, IEEE Computer, IEEE Network, IEEE Communications Magazine - Network Testing Series, IEEE Wireless Communications, IEEE Communications Surveys and Tutorials, IEEE Communications Letters, Computer Communications, Computer Networks, Journal of Network and Computer Applications, and IEICE Transactions on Information and Systems.* He has guest edited several Special Issues in IEEE journals and magazines, and co-chaired symposia at IEEE Globecom’13 and IEEE ICC’15. He published a textbook, *Computer Networks: An Open Source Approach* (www.mhhe.com/lin), with Ren-Hung Hwang and Fred Baker (McGraw-Hill, 2011). It is the first text that interleaves open source implementation examples with protocol design descriptions to bridge the gap between design and implementation.