CNS Research Review Agenda, Thursday, May 9, 2024

8:30 a.m. **Breakfast** [Computer Science and Engineering building at UC San Diego, Room 1242]

9:00 a.m. **Welcome and Round Table Introductions**
Stefan Savage and George Porter, CNS Co-Directors and CSE Professors

9:15 a.m. **Pat Pannuto - CNS Faculty/ Session One Chair**

- **Addressing Healthcare Ransomware Attacks (15)**
  Presenter: Stefan Savage, CNS/CSE Faculty
- **Unfiltered Measuring Cloud-based Email Filtering Bypasses (15)**
  Presenter: Sumanth Rao, CNS/CSE Graduate Student (Stefan Savage and Geoff Voelker)
- **Reverse Engineering + Securing an Insulin Pump (15)**
  Presenter: Alex Bellon, CNS/CSE Graduate Student (Deian Stefan and Pat Pannuto)
- **Using Honeybuckets to Characterize Cloud Storage Scanning in the Wild (15)**
  Presenter: Katherine Izhikevich, CNS/CSE Graduate Student (Geoff Voelker, Stefan Savage)
- **Towards High Interaction Chatbot Honeypots For Online Fraud (15)**
  Presenter: Daniel Spokoyny, CNS/CSE Graduate Student (Taylor Berg-Kirkpatrick, Stefan Savage and Geoff Voelker)

10:30 a.m. **Break and Informal Interaction (15 minutes)**

10:45 a.m. **Amy Ousterhout - CNS Faculty/ Session Two Chair**

- **Preemptive Userspace Scheduling with User Interrupts (15)**
  Presenter: Linsong Guo, CNS/CSE Graduate Student (Amy Ousterhout)
- **Mira: A Progam-Behavior-Guided Far Memory System (30)**
  Presenter: Zhiyuan Guo, CNS/CSE Graduate Student (Yiying Zhang)
- **Disaggregated Datastructures (30)**
  Presenter: Stewart Grant, CNS/CSE Graduate Student (Alex Snoeren)

12:00 p.m. **Group Photograph** [Bear Sculpture outside of Computer Science and Engineering]

12:15 p.m. **Lunch** [CSE Room 1202]

1:45 p.m. **Geoff Voelker - CNS Faculty/ Session Three Chair**

- **The Double-Edged Sword: Identifying Authentication Pages and their Fingerprinting Behavior (30)**
  Presenter: Alisha Ukani, CNS/CSE Graduate Student (Alex Snoeren)
- **The Effect of the Network in Cutting Carbon for Geo-Shifted Workloads (30)**
  Presenter: Yibo Guo, CNS/CSE Graduate Student (George Porter)
- **Stateful Least Privilege Authorization for the Cloud (30)**
  Presenter: Luoxi Meng, CNS/CSE Graduate Student (Earlence Fernandes)

3:15 p.m. **Break and Informal Interaction (30 minutes)**

CONTINUE TO PAGE 2

CNS
CENTER FOR NETWORKED SYSTEMS

Day One / CNS Research Review Agenda, Thursday, May 9, 2024

3:45 p.m. **Stefan Savage - CNS Faculty/ Session Four Chair**

- **Laurel: Automatic Repair of Dafny Proofs (15)**
  Presenter: Eric Mugnier, CNS/CSE Graduate Student (Yuanyuan Zhou)
- **Cachet: Trustworthy Just-In-Time Compilers with Symbolic Meta-Execution (30)**
  Presenter: Michael Smith, CNS/CSE Graduate Student (Deian Stefan)
- **Attacker-in-the-Middle Threats on Commercial and Military Aircraft (30)**
  Presenter: Aaron Schulman, CNS/CSE Faculty

5:00 p.m. **Lightning Talks – Graduate Students and Industry Representatives (30 minutes)**

6:00 p.m. **Graduate Student Poster Session and Dinner Reception** [15th Floor at the Village]

Day Two / CNS Research Review Agenda, Friday, May 10, 2024

9:00 a.m. **Breakfast** [Computer Science and Engineering building at UC San Diego, Room 1242]

9:30 a.m. **Alex Snoeren - CNS Faculty/ Session Chair**

- **Virtualizing Programmable Switches using Active Networking (30)**
  Presenter: Rajdeep Das, CNS/ CSE Graduate Student (Alex Snoeren)
- **Low-carbon Computing from Recovered Hardware (15)**
  Presenter: Jennifer Switzer, CNS/CSE Graduate Student (Pat Pannuto and Ryan Kastner)
- **RAN, PAN, or LPWAN: Towards the Future of Backhauling Data for Mobile Devices In the Wild (15)**
  Presenter: Alex Yen, CNS/CSE Graduate Student (Pat Pannuto)
- **Efficient Serving of Augmented Large Language Models (30)**
  Presenter: Yiying Zhang, CNS/CSE Faculty
- **IRRegularities in the Internet Routing Registry (15)**
  Presenter: Ben Du, CNS/CSE Graduate Student (Alex Snoeren and kc claffy)
- **Title TBD (30)**
  Presenter: TBD

11:45 a.m. **Open Floor** [Feedback from company representatives]

12:00 p.m. **Lunch and CNS Research Review Conclusion** [CSE Room 1202]

CNS Research Review Conclusion

- See abstracts on next page -

Abstracts in Order of Appearance:

Addressing Healthcare Ransomware Attacks - Stefan Savage

Ransomware attacks against healthcare delivery organizations have reach epidemic proportions with at least 460 such attacks against US HDOs in 2023 alone. Moreover, the impact of attacks in the healthcare setting is not a mere financial diaster, but can also have direct and quantifiable impact on the quality of care delivered. In this talk, I'll briefly summarize a new ARPA-H funded effort between members of CNS and UCSD Health focused on the unique clinical impacts of healthcare ransomware and how to best provide resilience.

Unfiltered Measuring Cloud-based Email Filtering Bypasses - **Sumanth Rao**

Email filtering is a popular mechanism to filter email threats, configured by an organization to filter inbound email. However, we show that a vast majority of such configurations are improperly secured and can be bypassed as a result, allowing for an attacker to send email effectively unfiltered.

Reverse Engineering + Securing an Insulin Pump - Alex Bellon

Personal medical devices, especially those that are physically attached to individuals at all times, have become more and more technologically capable in recent years. While these advances allow for better quality of life for users, new features can also pose potential security problems, which is especially important to avoid when a vulnerability could lead to dangerous effects on the user's health. We specifically look at a popular and feature-rich insulin pump, evaluating its current software for security issues with the ultimate goal of ensuring safe and correct execution of the pump. We present our work so far reverse engineering the pump, the initial security weaknesses we have found, and our plans for future improvement.

Using Honeybuckets to Characterize Cloud Storage Scanning in the Wild - Katherine Izhikevich

In this work, we analyze to what extent actors target poorly-secured cloud storage buckets for attack. We deployed hundreds of AWS S3 honeybuckets with different names and content to lure and measure different scanning strategies. Actors exhibited clear preferences for scanning buckets that appeared to belong to organizations, especially commercial entities in the technology sector with a vulnerability disclosure program. Most alarmingly, we recorded multiple instances in which malicious actors downloaded, read, and understood a document from our honeybucket, leading them to attempt to gain unauthorized server access.

Towards High Interaction Chatbot Honeypots For Online Fraud - Daniel Spokoyny

In this work, we propose using LLM chatbots as fake victim honeypots to understand conversational fraud. Instead of relying on reactive reporting from victims of this fraud or requiring humans to go along with scams by posing as victims, we show that LLM chatbots can be fine-tuned to operate as convincing stand-ins for human honeypots in order to later collect an automated conversational fraud dataset. Using human evaluation we show that our fine-tuned model significantly improves over baselines and the resulting conversations closely match in quality to the underlying self-instructed transcripts. Finally, we will describe our preliminary progress to build-out infrastructure required to interact with fraudsters.

Preemptive Userspace Scheduling with User Interrupts – Linsong Guo

Preemptive scheduling in datacenter systems is currently underutilized since the existing preemption mechanisms have high and unpredictable overheads. Intel's new feature, user interrupts, offers a promising solution by enabling low-overhead preemption entirely in userspace. This work investigates if user interrupts are suitable for micro-scale userspace preemption and leverage user interrupts to build a user-level scheduler.

Mira: A Progam-Behavior-Guided Far Memory System - Zhiyuan Guo
Far memory, where memory accesses are non-local, has become more popular in recent years as a solution to expand memory size and avoid memory stranding. Existing far memory system struggles to get a balance point between programmability and performance. We propose a new far-memory approach by automatically inferring program behavior and efficiently utilizing it to improve application performance. We build Mira, utilizes program analysis results, profiled execution information, and system environments together to guide code compilation and system configurations for far memory. Mira outperforms existing far memory systems up to 18 time.
The Effect of the Network in Cutting Carbon for Geo-Shifted Workloads - Yibo Guo
In this work, we investigate how we can reduce carbon impact of cloud computing via geographical workload migration. In particular, we propose an approach to geographic workload migration that uses high-fidelity maps of physical Internet infrastructure to better estimate the carbon costs of WAN transfers. **(Research funded by CISCO.)**

The Double-Edged Sword: Identifying Authentication Pages and their Fingerprinting Behavior - Alisha Ukani
Browser fingerprinting is often associated with cross-site user tracking. However, less is publicly known about its uses to enhance online safety, where it can provide an additional security layer against service abuses (e.g., in combination with CAPTCHAs) or during user authentication. Our work explores the question of whether fingerprinting is used for tracking or for improving user security. We measure the fingerprinting behavior of the authentication pages for the top 100K websites and find that fingerprinting is often used for both purposes—sometimes simultaneously. This work highlights the need for more privacy-preserving tools to improve user security.

Stateful Least Privilege Authorization for the Cloud - Luoxi Meng
Cloud service developers architect their permission model to give out more access to user resources than what client apps need. For example, when Zoom integrates with Google Calendar, Zoom obtains a bearer token --- a credential that grants broad access to user data on the server. Widely-used authorization protocols like OAuth result in this undesirable situation because they do not provide developers of client apps and servers the tools to request and enforce minimal access. In the status quo, these overprivileged credentials are vulnerable to abuse when stolen or leaked. We introduce an authorization framework that enables creating and using bearer tokens that are least privileged. Our core insight is that the client app developer always knows their minimum privilege requirements when requesting access to user resources on a server. Our framework allows client app developers to write small programs in WebAssembly that customize and attenuate the privilege of OAuth-like bearer tokens. The server executes these programs to enforce that requests are least privileged. Building on this primary mechanism, we introduce a new class of stateful least privilege policies --- authorization rules that can depend on a log of actions a client has taken on a server. We instantiate our authorization model for the popular OAuth protocol. Using open source client apps, we show how they can reduce their privilege using a variety of stateful policies enabled by our work.

Laurel: Automatic Repair of Dafny Proofs - Eric Mugnier
Dafny is a popular verification language, which automates proofs by outsourcing them to an SMT solver. This automation is not perfect, however, and the solver often requires guidance in the form of helper assertions, creating a burden for the proof engineer. In this paper, we propose Laurel, a tool that uses large language models (LLMs) to automatically generate helper assertions for Dafny programs. To improve the success rate of LLMs in this task, we design two domain-specific prompting techniques. First, we help the LLM determine the location of the missing assertion by analyzing the verifier's error message and inserting an assertion placeholder at that location. Second, we provide the LLM with example assertions from the same codebase, which we select based on a new lemma similarity metric. We evaluate our techniques on a

dataset of helper assertions we extracted from three real-world Dafny codebases and show that it can synthesize more than half of these assertions.

Cachet: Trustworthy Just-In-Time Compilers with Symbolic Meta-Execution - Michael Smith
Just-in-time (JIT) compilers face a tough task: chewing up potentially-untrusted input and spitting out machine code that's at once fast, correct, and secure, working close-to-the-metal with low-level primitives in a tight time budget. Developers juggle complex invariants which generated code must respect, which can easily shatter the system's security model when broken. Cachet, our domain-specific language for JIT implementation, makes these invariants explicit and statically-verified. Our toolchain compiles Cachet code both to the SMT-solvable Boogie verification language and to C++ suitable for embedding in host applications and language runtimes. We evaluate Cachet by re-implementing and verifying components of Firefox's JavaScript JIT.

Attacker-in-the-Middle Threats on Commercial and Military Aircraft - Aaron Schulman
In this talk, I will describe a new kind of vulnerability where an attacker can attach a rogue device to a data bus and achieve attacker-in-the-middle capabilities without physically being in the middle of its wires. This is a threat to critical equipment operating in the field which often has outwardly accessible receptacles connecting to sensitive data buses. I will present the results of our study of whether this attack is feasible on two popular types of data buses used to control airplanes and boats. I will also provide a real-world case study of a popular commercial aircraft that may be vulnerable to the attack through an unused maintenance outlet. In our testing, we found that by simply plugging a device into this outlet, an attacker can covertly modify the flight plan that the pilots load into the aircraft's flight computer without showing the change on one of the primary displays in the cockpit.

Virtualizing Programmable Switches using Active Networking - Rajdeep Das
We consider the problem of enabling user-defined functionality and resource sharing on programmable switch hardware using active networking. We present approaches to programming, hitless provisioning, software isolation and memory management on PISA programmable switches. **(Research funded by CISCO.)**

Low-carbon computing from recovered hardware - Jennifer Switzer
Sustainable computing efforts have traditionally focused on runtime efficiency. However, a significant fraction of the carbon emissions associated with computing is incurred not during use, but rather manufacture. These manufacturing emissions are responsible for 40% of the lifetime carbon footprint for server-class hardware. To reduce manufacturing emissions, we propose the repurposing of unwanted consumer-class devices as general purpose compute. We find that repurposed devices can provide a computing platform that is several times more carbon-efficient than the alternative of manufacturing new hardware.

IRRegularities in the Internet Routing Registry – Ben Du
Despite being the most widely-used database for routing security, the Internet Routing Registry's (IRR) lack of strict validation standards and limited coordination among database providers leads to inaccuracies and security risks. We conducted a 1.5-year longitudinal analysis of the IRR and identified 6,373 deemed potentially suspicious records. Our findings underscore the need for enhanced validation processes to address the misuse of IRR records for attacks on the Internet routing system.