Colloquium: Security & Privacy

3. – 7. February 2025
Showroom (DHEG136E) | Sandgasse 36, Erdgeschoß

It is a pleasure to invite you to the colloquium for our Professorship in Security & Privacy at Graz University of Technology. The public part will be a short educational presentation at Bachelor's level 3rd year in Computer Science on topic Digital Signature and Applications, a scientific talk (titles below), and a discussion with the audience.

Apruzzese Giovanni

3. February 2025 | 08:30 | Showroom (DHEG136E) | Sandgasse 36, Erdgeschoß

Title: "The many faces of AI in the Phishing-website landscape"

Abstract: Phishing websites are everywhere. This fact may come at a surprise when considering the thousands of papers proposing artificial-intelligence (AI) techniques to counter this threat. Some of these techniques "work", i.e., they can reliably detect phishing websites—which is clearly an encouraging result. However, many "state-of-the-art" AI methods can also be trivially fooled with little effort by naive attackers—which is clearly a disheartening result. Lastly, AI methods can also be offensively used by attackers to circumvent AI-based detectors—which is clearly a worrying result.

In this talk, I will explore these three complementary classes of results, each denoting a different "face" of AI. Specifically, I will explain how AI can be used to catch phish. Then, I will show how to trivially evade these AI-based methods with simple modifications that anyone could do. Finally, I will reveal more sophisticated (but still affordable) ways to maliciously use AI tools to circumvent phishing detectors powered by AI. During this journey I will also emphasize the role of the end-user: ultimately, a phishing website must deceive a human—not an AI.

Abdelnabi Sahar

3. February 2025 | 14:00 | Showroom (DHEG136E) | Sandgasse 36, Erdgeschoß

Title: "Towards Aligned, Interpretable, and Steerable Safe AI Agents"

Abstract: AI models are becoming more ubiquitous in our everyday lives. We now have so many real-world AI-integrated products and applications that are used by millions of users on a daily basis. While these applications can enhance utility and automate tasks, there are many challenges in ensuring their reliability and trustworthiness. In this talk, I will briefly discuss my work that was the first to uncover the indirect prompt injection vulnerability in LLM-integrated applications, which is one of the most pressing security and safety risks to LLM-integrated applications and largely acknowledged by large tech companies, practitioners, and researchers. I will then talk about our follow-up work to detect this threat based on models' internal states, providing significant improvements over text-based classification methods. I will discuss our ongoing work, via a public competition, to build community-based adaptive attacks and benchmarks for indirect prompt injection. Beyond single models, AI agents are the future of automating many workflows and creating cooperative agentic systems that communicate, deliberate, and solve tasks. I will finally discuss challenges in evaluating and securing multi-agent workflows and multi-agent manipulation risks. I will conclude with sharing a future vision to ensure Aligned, Interpretable, and Steerable Safe AI Agents.

Fassl Matthias

4. February 2025 | 08:30 | Showroom (DHEG136E)| Sandgasse 36, Erdgeschoß

Title: "Designing for Social Cybersecurity and Safety"

Abstract: Since cryptography became more accessible in the 1960s, governments have tried to limit access to it in several crypto wars. These discussions pitch privacy and safety, both very valuable, against each other, saying that one undermines the other. However, safety comes not only from police investigations and increased surveillance but also from deploying hard-to-abuse tools and providing support when needed. Careful design can help achieve these goals. Making technology harder to abuse and harm others and offering a technology-mediated version of the social support we know from our everyday lives. Based on my research on the role of social norms in security and privacy behavior and designing safety mechanisms against intimate-image-based abuse, I will present my vision of policy-oriented HCI research in cybersecurity that supports safety without surveillance.

Wei Miranda

4. February 2025 | 14:00 | Showroom (DHEG136E)| Sandgasse 36, Erdgeschoß

Title: "Leveraging Sociotechnical Security and Privacy to Address Online Abuse "

Abstract: The prevalence and severity of online abuse are on the rise, from toxic content on social media to image-based sexual abuse, as new technologies are weaponized by people who do harm. Further, this abuse disproportionately harms people already marginalized in society, creating unacceptable disparities in safety and reinforcing oppression. Working in the areas of both computer security and privacy (S&P) and human-computer interaction (HCI), I address online abuse as the next frontier of S&P challenges. In this talk, I discuss my approach to sociotechnical threat modeling that (1) characterizes emerging S&P threats in digital safety, with particular attention to the technical and societal factors at play, (2) evaluates the existing support for online abuse, taking an ecosystem-level perspective, and (3) develops conceptual tools that bridge S&P and HCI towards societally informed S&P research. I conclude by outlining how sociotechnical security and privacy can work towards a world where all people using technology feel safe and connected.

de Melo Branco Pedro

5. February 2025 | 08:30 | Showroom (DHEG136E)| Sandgasse 36, Erdgeschoß

Title: "Succinct Cryptography"

Abstract: In this talk, we introduce the field of succinct cryptography. The major goals of this area are to i) Construct cryptographic protocols with good communication complexity, ii) Diversify the hardness assumptions from which we can build them, and iii) Aim for the best security notion possible. As a concrete example we present a series of results on communication-efficient oblivious transfer (OT), a cryptographic primitive which is at the basis of multi-party computation protocols. These results achieve OT protocols with optimal rate, achieving different notions of security.

Schloegel Moritz

5. February 2025 | 14:00 | Showroom (DHEG136E)| Sandgasse 36, Erdgeschoß

Title: "Have we solved fuzzing yet? "

Abstract: Over the past decade, fuzzing has established itself as one of the most effective bug-finding techniques. Spurred by the introduction of coverage feedback, fuzzing research has experienced a renaissance: Hundreds of papers promised to improve almost all its aspects, boosting the fuzzers' effectiveness and continuously pushing the boundaries of their applicability. Yet, many techniques never found widespread adoption, and anecdotal evidence points to biased or flawed evaluations, raising the question of whether we can reproduce many of these proposed improvements. At the same time, some targets, such as web servers, elude almost all fuzz testing efforts, even though they expose a vast attack surface -- challenging our notion that fuzzing is widely scalable and usable. In this talk, we will systematically review the reproducibility of fuzzing, study what can go wrong, and discuss how we can do better. Then, we will discuss a novel approach to test critical targets, including web servers, that elude existing approaches. Grounded in fault injection, our fuzzer exploits the domain knowledge encoded in already available programs to test challenging targets effectively. Concluding this talk, we will explore open problems that future work needs to address to facilitate mainstream adoption of fuzzing.

Utz Christine

6. February 2025 | 14:00 | Showroom (DHEG136E)| Sandgasse 36, Erdgeschoß

Title: Beyond Notice and Consent: Fostering Online Privacy in the Age of AI
Abstract: The constant emergence of new data breaches and the fines imposed for violations of data protection laws demonstrate that the providers of digital services, software, and devices often fail to respect their users' right to privacy of their personal information. They mostly rely on the principle of "notice and consent" and prompt users to agree to often extensive data collection and sharing, and many harmful data practices remain undeclared and invisible to users. A more holistic approach is privacy by design, which considers user privacy throughout all stages of the development process. While this principle is enshrined in Article 25 of the GDPR and privacy-friendly alternatives to popular technology do exist, they rarely find widespread adoption.
This talk investigates this phenomenon, possible reasons, and ways to move forward. It demonstrates insufficient privacy protections in digital ecosystems at the example of HbbTV (Hybrid Broadcast Broadband TV), an evolving TV standard that combines linear TV programming with interactive Web content, thus creating opportunities for user tracking. Obstacles in the adoption of privacy by design on the Web are investigated via an online survey that finds website operators frequently unaware of their legal obligations and the data processing practices of widely employed development tools. Fostering more holistic approaches to protecting user privacy is a complex issue but increasingly important in an AI-driven world, where potential future uses of user data and the inferences drawn from it are becoming nearly impossible to predict.

Krstic Srdan

7. February 2025 | 08:30 | Showroom (DHEG136E)| Sandgasse 36, Erdgeschoß

Title: “Runtime Enforcement for Privacy”
Abstract: Data protection regulations in many countries require IT systems to implement baseline privacy requirements. These requirements are challenging to precisely state, let alone implement, as system developers must address them consistently and in a cross-cutting manner.
Given a policy describing the desirable behavior of some target system, runtime enforcement is a general technique that constructs another system, called an enforcer, that observes and actively controls the behavior of the target system by modifying its actions to ensure policy compliance.
In this talk, I will (1) provide a brief overview of runtime enforcement, (2) show how common privacy requirements (e.g., from the GDPR) can be formulated as policies, and (3) demonstrate how runtime enforcement can be used to ensure policy compliance. The main takeaway from my talk is that one should specify privacy requirements at the target system's design phase and then rely on runtime enforcement, rather than coding the requirements directly during the system's implementation phase. Our empirical evidence conclusively suggests this for a similar question about security

requirements.

Roth Sebastian

7. February 2025 | 14:00 | Showroom (DHEG136E)| Sandgasse 36, Erdgeschoß

Title: “A Comprehensive Approach for System Security”
Abstract: We are using the Web on a daily basis for banking, communication, education, collaboration, entertainment, and more. Due to this importance, the Web is one of the main targets of attacks, most prominently Cross-Site Scripting (XSS). Luckily a correctly crafted Content Security Policy (CSP) is capable of effectively mitigating the effect of those attacks. Throughout my research journey so far we have shown that the vast majority of all policies in the wild are, and have always been, trivially bypassable. To uncover the root causes behind the omnipresent misconfiguration of CSP, we conducted a qualitative study involving 12 real-world Web developers. By using methods from social science such as semi-structured interviews, enriched with a drawing task and a programming task, we were able to identify the participants' motivations, roadblocks for secure deployment, and strategies used to create a CSP. Due to the infeasibility of breaking changes in an already widely deployed mechanism, we used a similar methodology to also shed light on usability issues of a novel, not yet widely deployed, approach to defend against XSS, namely Truted Types. Based on the here detected issues we are currently in the process of conducting a participatory design approach together with Website operators, Web developers, security experts, and browser vendors, to come up with a mechanism that is both: usable and secure.