Collaborative Research: CPS: Medium: Enabling Data-Driven Security and Safety Analyses for Cyber-Physical Systems

Adwait Nadkarni (PI), William & Mary; Denys Poshyvanyk (Co-PI), William & Mary; Kevin Moran (Co-PI), University of Central Florida (Co-PI)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=2132281

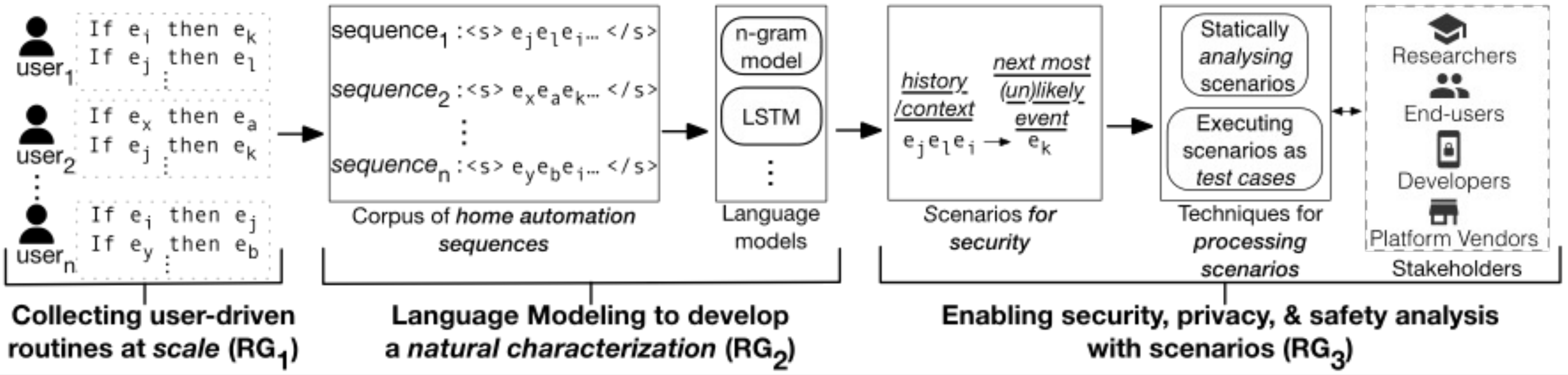
- **Smart home products are extremely popular** due the tremendous convenience offered through home automation
- Due to bridging cyber-physical gap, **home automation signifies a widening of the attack surface** of the home
- **Existing research lack of a realistic characterization** of home automation usage as deployed by end-users
- **User-driven automation**, a natural expression of user-requirements, **can be modeled using statistical language models** (LMs) for security research

Challenges

- Collecting and Tokenizing User-driven Routines at Scale requires **maintaining naturalness**
- User Routines need to be be modeled as Home Automation Sequences while **maintaining temporal and semantic relationships between different events and routines**.
- Generating effective scenarios with security flavors, i.e., unnatural but interesting scenarios, which can enable security, privacy, and safety analysis

Scientific Impact

- Laying the groundwork for adapting LM techniques for modeling trigger-action home automation
- Paradigm-shift in home-automation analysis and characterization for security
- Laying the groundwork for adapting existing LM techniques in the SE domain to model the trigger-action programs
- Enable stress-test for platform/device/app security engineers with a large and diverse array of natural scenarios as test cases in realistic circumstances



```
user1   If ei then ek
        If ej then el
user2   If ex then ea
        If ej then ek
usern   If ei then ej
        If ey then eb
```

$sequence_1$: <s> $e_j e_l e_i$… </s>

$sequence_2$: <s> $e_x e_a e_k$… </s>

$sequence_n$: <s> $e_y e_b e_i$… </s>

Corpus of *home automation sequences*

n-gram model

LSTM

Language models

history /context $e_j e_l e_i$ → *next most (un)likely event* e_k

Scenarios *for security*

Statically *analysing* scenarios

Executing scenarios as *test cases*

Techniques for *processing scenarios*

Researchers

End-users

Developers

Platform Vendors

Stakeholders

Collecting user-driven routines at *scale* (RG_1)

Language Modeling to develop a *natural characterization* (RG_2)

Enabling security, privacy, & safety analysis with scenarios (RG_3)

- Facilitate timely discovery and patching of vulnerabilities in individual products and third-party integrations before released to users
- Enabling the understanding of the security and privacy implications of their workflows for the users
- More secure, private, and useful home automation, thus enhancing the overall user experience, and contributing to the adoption of secure smart home technology

- Graduate and Undergraduate level projects in classes across W&M, and collaborative work with IIT, DU to enhance knowledge and skills
- An educational tool that allows students to create meaningful smart home routines and execute them in realistic execution environments
- K-12 guest lectures and outreach events to involve underrepresented students in graduate research, such as the W&M CS Graduate Symposium

- A large-scale collection of thousands of user-driven routines, and realistic home automation event sequences, obtained from real users.
- Hundreds of security policies obtained from statically analyzing the realistic home automation scenarios.
- Vulnerabilities and gaps in real security tools for smart homes obtained from executing scenarios as test cases with the tools.