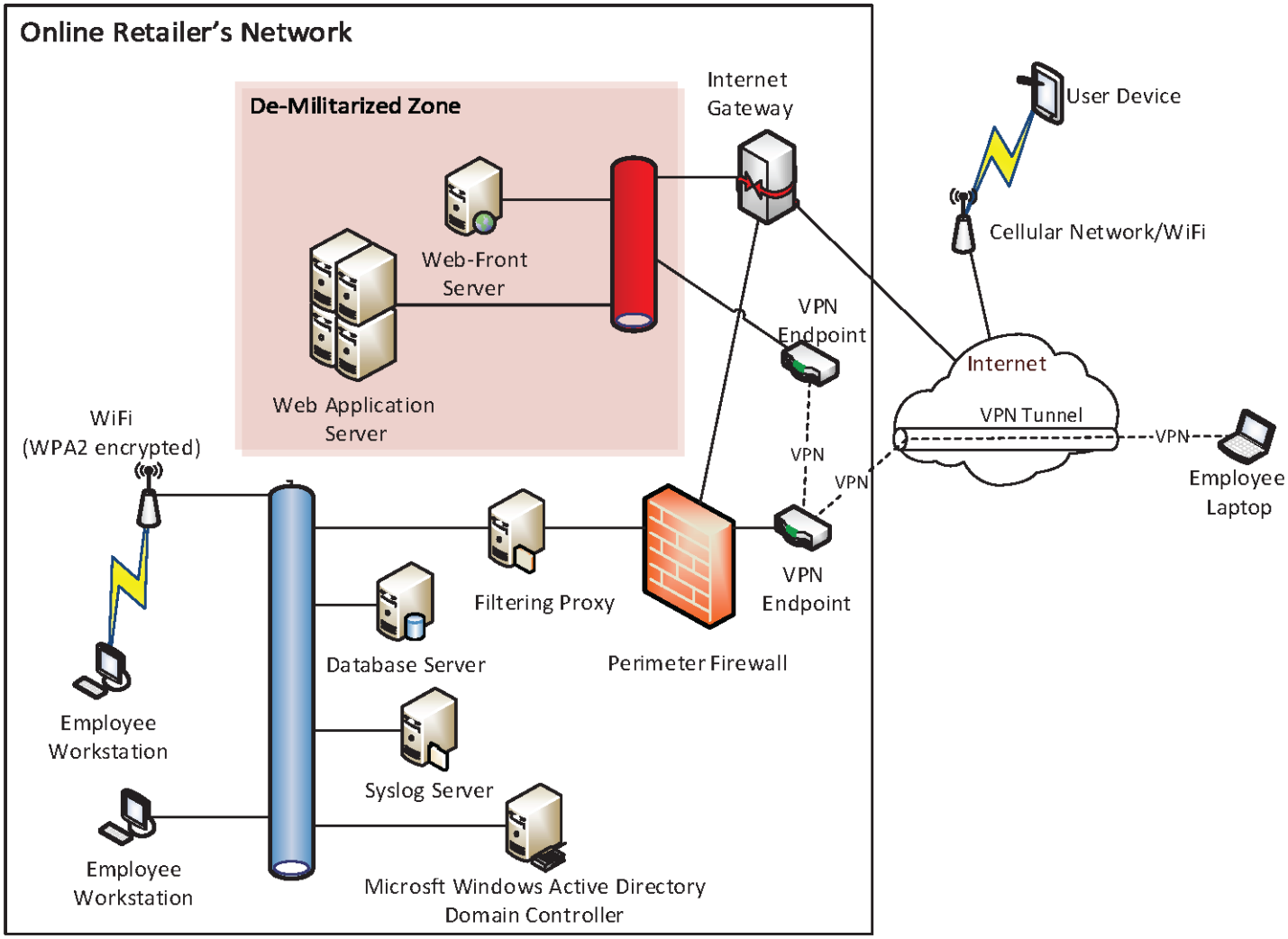
Carnegie Mellon University

Decision Support for Cybersecurity Risk Assessment

Hanan Hibshi & Travis D. Breaux
April 26, 2017

institute for SOFTWARE RESEARCH

Security in a Composable System



Security Checklists

- Security "*Best Practices",* for example:
  - NIST (National Institute of Standards and Technology) publications 800 series
  - OWASP (The Open Web Application Security Project)
- Does not consider context

(1) *AUTHENTICATOR MANAGEMENT | PASSWORD-BASED AUTHENTICATION*

The information system, for password-based authentication:

(a) **Enforces minimum password complexity of [*Assignment: organization-defined requirements for case sensitivity, number of characters, mix of upper-case letters, lower-case letters, numbers, and special characters, including minimum requirements for each type*];**

(b) **Enforces at least the following number of changed characters when new passwords are created: [*Assignment: organization-defined number*];**

(c) **Stores and transmits only cryptographically-protected passwords;**

(d) **Enforces password minimum and maximum lifetime restrictions of [*Assignment: organization-defined numbers for lifetime minimum, lifetime maximum*];**

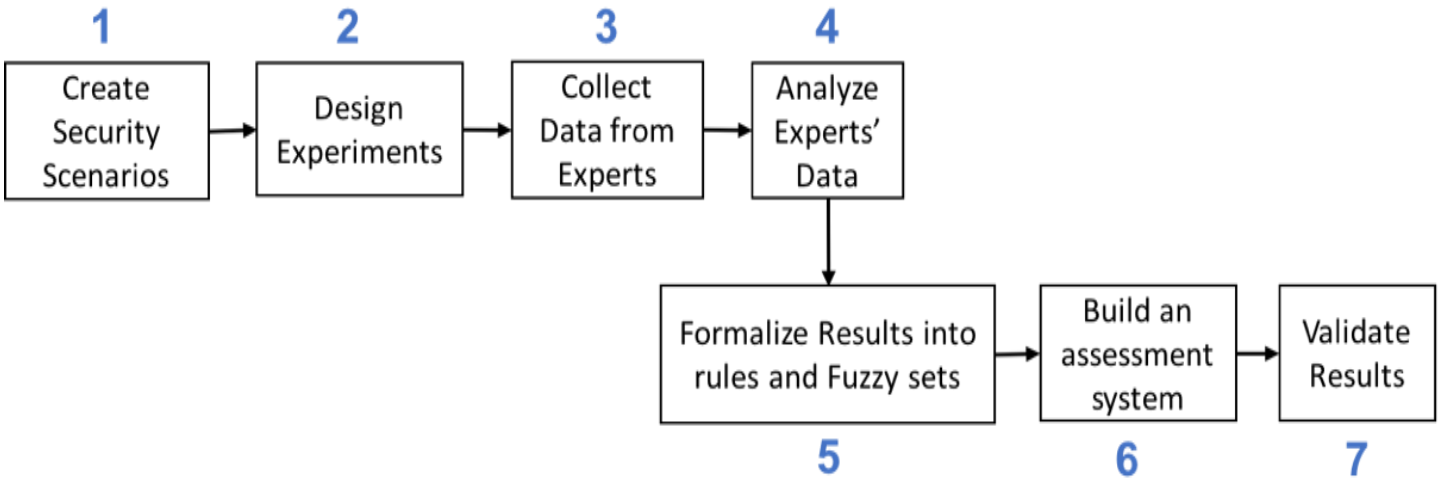
(e) **Prohibits password reuse for [*Assignment: organization-defined number*] generations; and**

(f) **Allows the use of a temporary password for system logons with an immediate change to a permanent password.**

Security Challenges

- There is a shortage in experts, in 2014: *U.S. Bureau of Labor statistics, Cisco
  - 82,900 information security analysts in the U.S.
  - median earn = $89,000 a year
  - 1M shortage in security professionals
  - unfilled positions ≈ 209,000
- By 2018, 53% growth in demand for experts is expected
- We need decision-support
- We focus on 3 challenges:
  - Varying level of *security expertise*
  - *Composition:* how security requirements work together
  - *Uncertainty:* present in security decisions

Our Overall Research Process



The Vignette Template

You are working on your laptop using **$NetworkType**. You are **$Transaction**. You are relying on a web browser to perform your task. The browser is already using **$Connection** for the session. To log in to the system and start your task, you will need to authenticate using a password that **$Password**. The system will **$Timer**.

The **$Threat** attack is a serious security concern. Please answer the following questions with regards to mitigating this threat.

Survey Mechanics

- Participant is asked to consent
- Presented with four MiTM scenarios varying in `$NetworkType`
- Asked to answer 10 security knowledge questions
  - Cryptography, network administration, systems, etc.
- Asked to answer some demographic questions
  - Job experience
  - Security training
- After 1-2 weeks, we asked participants to return and repeat the survey by for a different threat (Packet-Sniffing).

Sample Demographics

- Recruited from security classes mailing lists
  - CMU and NCSU
- 174 participants took the M-i-t-M survey
  - 116 returned and took the packet-sniffing
- 73% Males, 26% Females, 1% unreported
- Age groups: 18-24 (63%), 25-34 (33%), and 35+ (3%)
- 101 graduate students, 42 undergraduate students and 2 university professors.

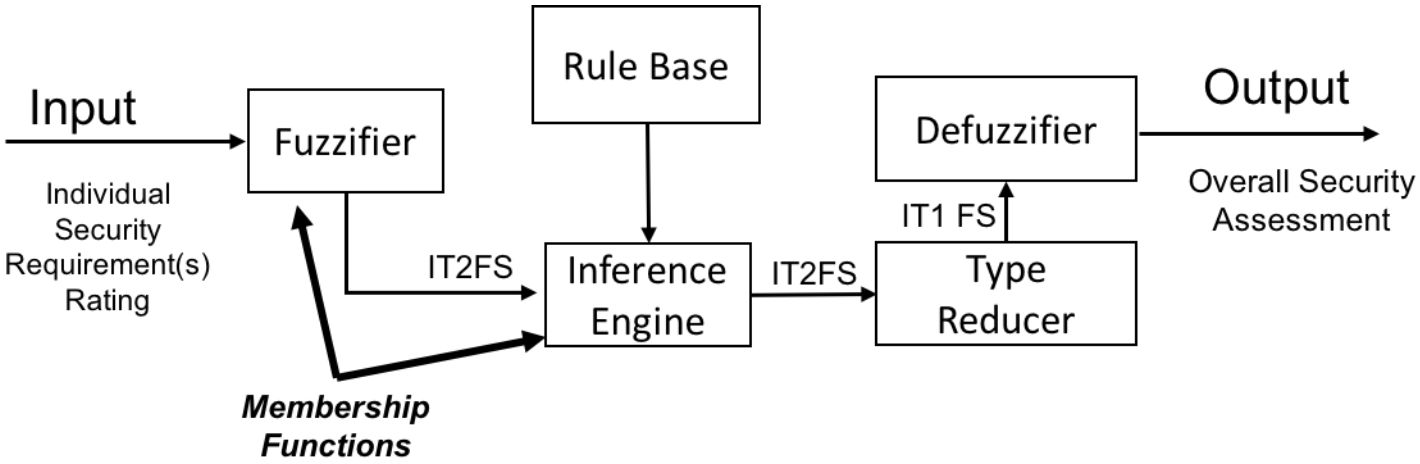
Interpreting Results into Rules

- Because network type takes priority over other requirements, rules R7- R81 are removed
- Example Rule:

$$R^1: IF\ \boldsymbol{NetworkType}\ is\ \boldsymbol{Inadequate}$$
$$THEN\ \boldsymbol{OverallRating}\ is\ \boldsymbol{Inadequate}$$

R #	Antecedents (IF)				Con. (THEN)
	Network	*SSL*	*Password*	*Timer*	*Overall*
R1	I				I
R2	A	I			I
R3	A		I		I
R4	A			I	I
R5	A	A	A	A	A
R6	E	E	E	E	E

Security Assessment System



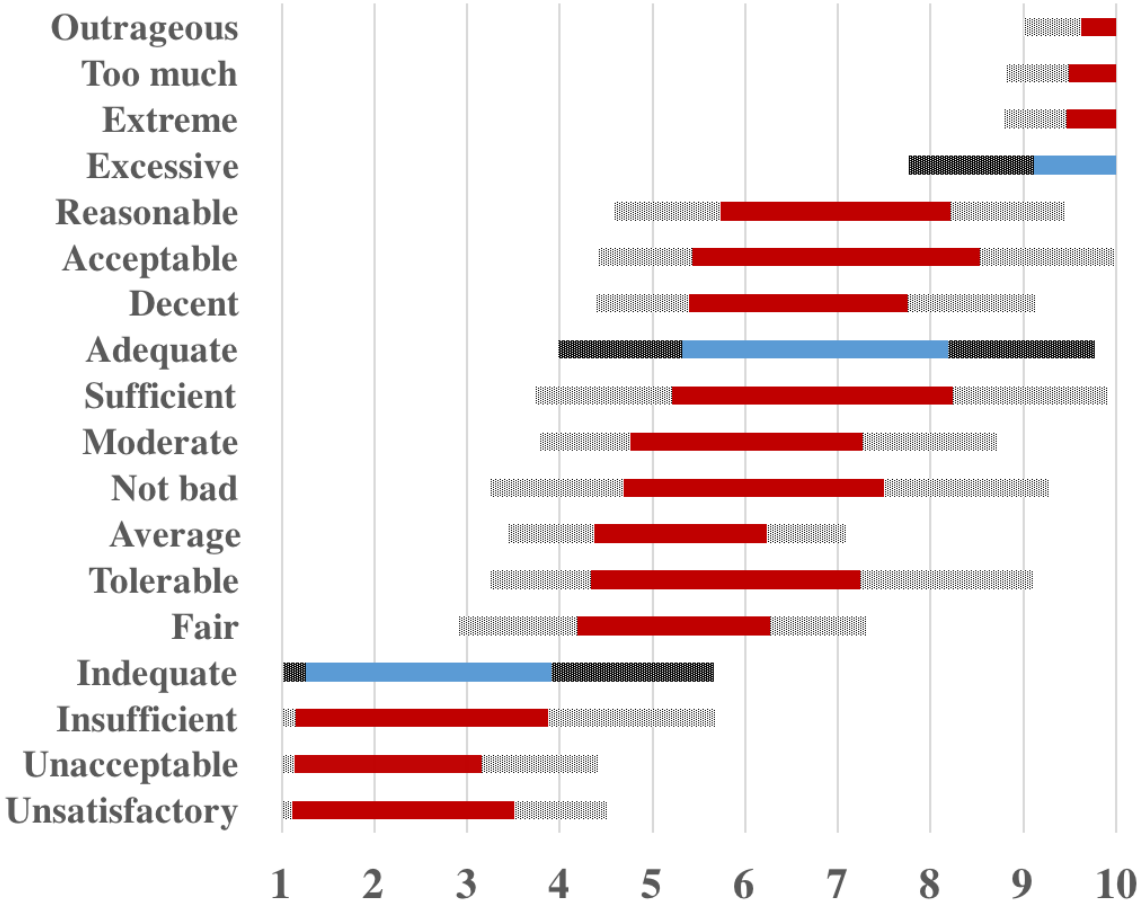
- Interval Type 2 Fuzzy Logic System (IT2FLS)
- Rule-based system
- Accounts for interpersonal and intrapersonal uncertainty

Adequacy Linguistic Labels for Security

- Focus group to select initial sets:
  - Inadequate, Adequate, and Excessive
- Expanded the set based on thesaurus
- Empirically evaluate 17 words [1]
  - Used 4 scenarios with different skewness or bias
    - Waiting for a bus
    - Distance to parking
    - Meal portion at a restaurant
    - Amount of privacy against government surveillance
  - Participants from Mechanical Turk, screened for English proficiency
    - Nelson-Denny English Test
- Assigning intervals for labels

[1] H. Hibshi and T. D. Breaux, “Evaluation of Linguistic Labels Used in Applications,” Tech. Rep, Carnegie Mellon Uni., 2016.

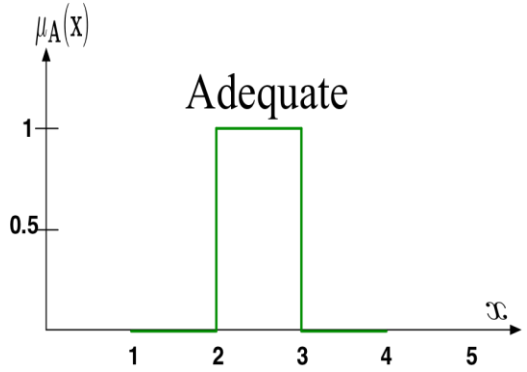
Fuzzy Sets for Adequacy Levels



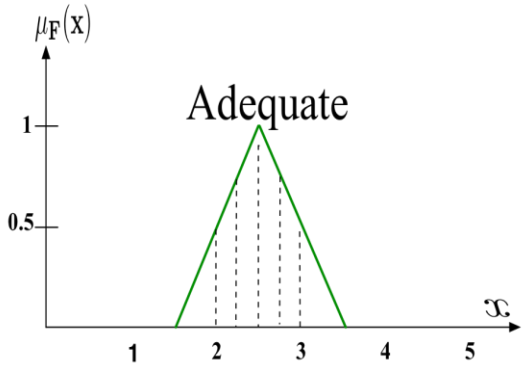
Fuzzy Sets and Membership Functions

- Fuzzy set theory expresses to what degree and element x belongs to a set.

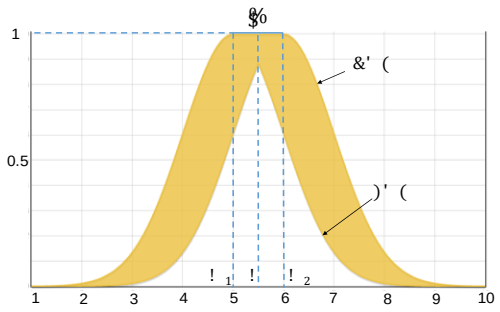
$$F = \{(x, \mu_F(x)) | \, x \in X\}$$



(a) Crisp sets

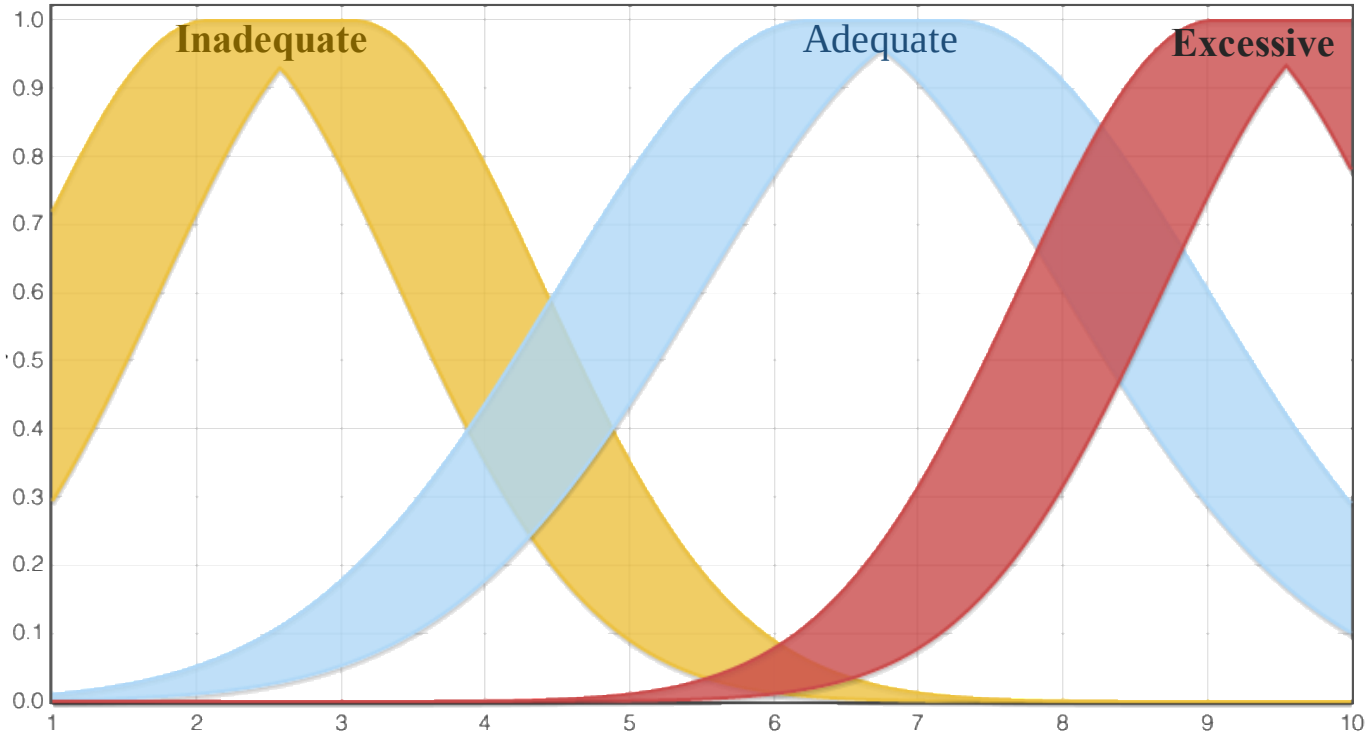


(b) Fuzzy sets



(c) Type-2 FOU constructed by blurring a Type-1 MF

Adequacy Membership Functions



Evaluation

- Interviewed 13 experts, 4 test scenarios each
- Disagreement: 19% of test scenarios
  - System assessment was more conservative then participants
- System demonstrated ability to augment human's shortfall in memory

Scenario			Total Participants	Agreement Ratio
Network (Wi-Fi)	*Pwd*	*Timer*		
Public unencrypted	Weak	None	5	4/5 (80%)
Public unencrypted	Weak	15-min	8	6/8 (75%)
Public unencrypted	Strong	None	8	6/8 (75%)
Public unencrypted	Strong	15-min	5	3/5 (60%)
VPN over encrypted	Weak	None	8	6/8 (75%)
VPN over encrypted	Weak	15-min	5	2/5 (40%)
VPN over encrypted	Strong	None	5	2/5 (40%)
VPN over encrypted	Strong	15-min	8	4/8 (50%)

Future Work

- Adapt the method for more multi-step attack vectors
- Enable mitigations recommendations to achieve higher overall security ratings
- Recruit industry experts
  - Already recruited around 80 experts from one conference
  - Four security domains:
    - Networking
    - Operating systems
    - Databases
    - Web applications

Questions?

- This research was funded by:
  - National Security Agency (Award #141333), and
  - Office of Naval Research (Award #N00244-16-1-0006)

- Thank you hhibshi@cmu.edu

- Hibshi, H., Breaux, T.D. and Broomell, S.B., 2015, August. Assessment of risk perception in security requirements composition. In Requirements Engineering Conference (RE), 2015 IEEE 23rd International (pp. 146-155). IEEE.
- Hibshi, H., Breaux, T.D. and Wagner, C., 2016. Improving security requirements adequacy: an interval type 2 fuzzy logic security assessment system.
- Hibshi, H., and Breaux, T.D., Reinforcing Security Requirements with Multifactor Quality Measurement . *Under submission*.

Vignette Questions

Overall, how would you assess the security of the system in the scenario above?

- ◯ **Inadequate** security measures that not enough to mitigate the threat
- ◯ **Excessive** security measures that exceeds the requirements to mitigate the threat
- ◯ **Adequate** security measures that is enough to mitigate the threat

Please rate the following variables based on their ability to mitigate the threat (Man-in-the-Middle attack):

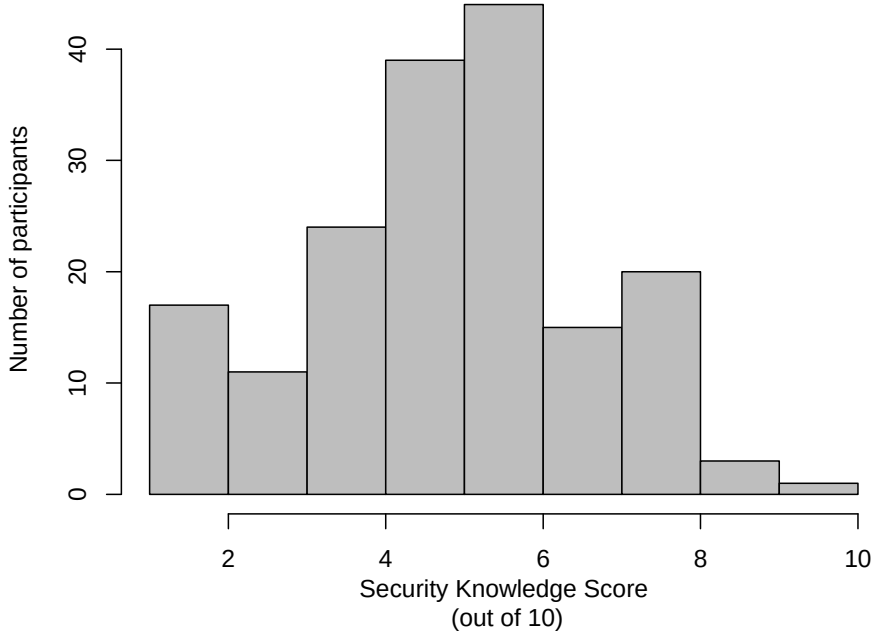
	Inadequate Mitigation 1	2	Adequate Mitigation 3	4	Excessive Mitigation 5
The **network** is employer's VPN that you connected to through public encrypted Wi-Fi	◯	◯	◯	◯	◯
The browser is using **SSL** for the session	◯	◯	◯	◯	◯
The **password** is at least 8 characters long	◯	◯	◯	◯	◯
The **timer** will automatically log you off the session after 15 minutes of inactivity	◯	◯	◯	◯	◯

Q: Please list down any **additional mitigations** that **will raise the security level of the scenario** above

Our Sample's Knowledge Effect

- Participants responses to security knowledge test
- `$Score` variable with values 0 -10
- Min=1; Max=10; Mean=5.2; Median= 5

- Participants with higher `$Score` gave lower ratings for: `$PasswordRating`, and `$TimerRating` in the presence of MiTM



Example of Security Questions

Why would an administrator set these firewall rules?

- `iptables -A OUTPUT -o eth0 -p tcp --dport 22 -m state --state NEW,ESTABLISHED -j ACCEPT`
- `iptables -A INPUT -i eth0 -p tcp --sport 22 -m state --state ESTABLISHED -j ACCEPT`

Which of the following is considered good encryption for files on your hard disk?

- SSL
- PGP
- SHA256
- AES

Power

- Mixed-effects (between and within subjects)
- Multi-level regression limits the biased covariance estimates
- 81% higher sample size than estimated

Threats to Validity

- Internal Validity
  - Randomized assignment to conditions
  - Learning and fatigue:
    - 20 min average survey time
    - One week span between two surveys
- External Validity
  - Target security experts
  - Bias: recruitment from US universities
- Construct validity
  - Defined rating levels
  - Tested the terms used in another Mturk survey of ~300 participants
- Power
  - Mixed-effects (between and within subjects)
  - Multi-level regression limits the biased covariance estimates

Choice of Experimental Method

- Traditional survey designs of direct questions or rating of a statement ➔ not sufficient
- We are studying human judgment
  - The context effect
  - The underlying components and their interaction
  - People are unaware of components
- Different experimental methods exist
  - Humans rely on their evaluation of factors they perceive in a situation to build a decision [1]
  - Scenario-based methods

[1] P. H. Rossi and S. L. Nock, *Measuring Social Judgments: The Factorial Survey Approach*. SAGE Publications, 1982