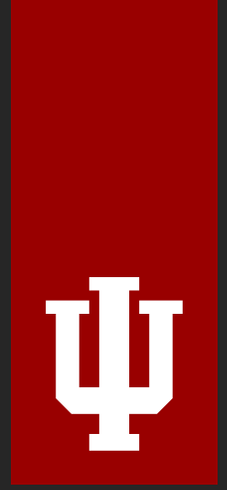


A Proof of Concept for Usability and Efficacy Evaluations as a Component of IETF Standards Using MUD

Vafa Andalibi, Jayati Dev

Indiana University Bloomington

Efficacy of Standards

DNSSEC

- DNS has two failure:
  - Name doesn’t exist
  - Connectivity error
- DNSSEC has two new failure cases:
  - Requestor under attack (or is it?)
  - Expired configuration

DNSSEC

- Problem?
  - Software handles these two new failure cases
  - It would get harder for users to “Click Through” certificate warnings
    - CURL had to change its API to handle this:
      - VERIFYHOST=0: it doesn't validate SSL certificates
      - VERIFYHOST=2: it verifies SSL certificates
      - VERIFYHOST=1: It checks to see if the certificate attests to any hostnames, and then accepts the certificate **no matter who presents it**

RPKI

- BGP routing isn’t secure
- Secure web browsing deployed PKI, but BGP route validation has not moved forward, why?
  - All networks would need to be embrace RPKI (and more)
  - Acceptability should have been considered

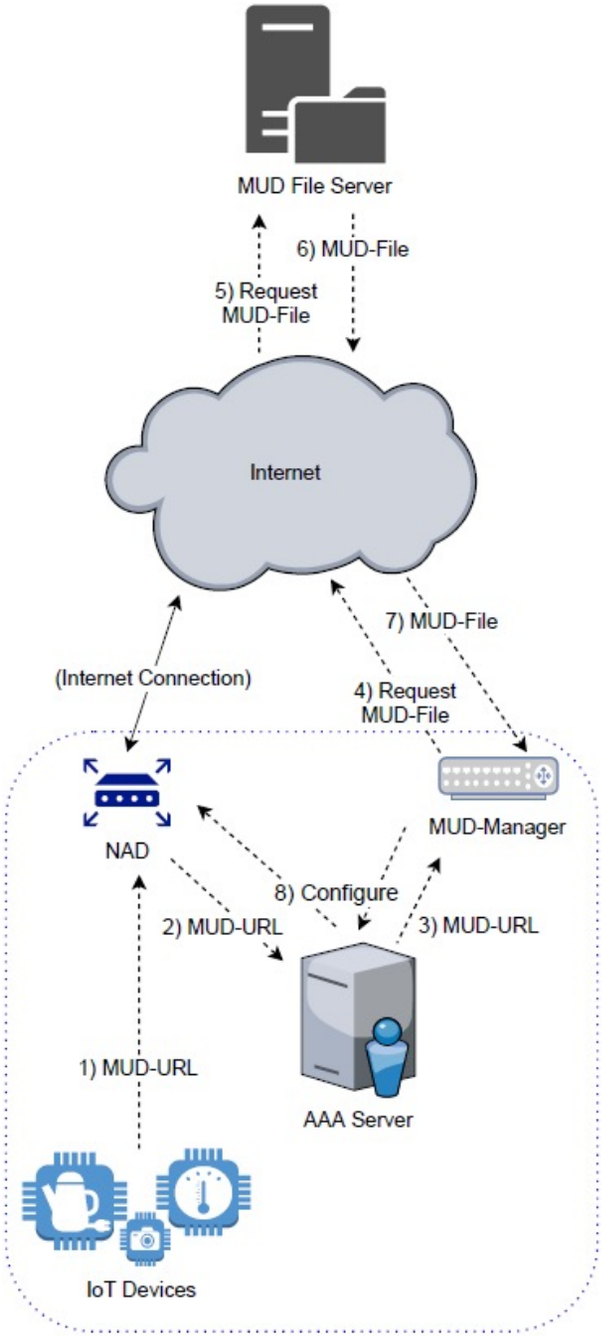
IPV6

- Intention: Just like IPv4, except with128-bit source and address fields
- Some Issues:
  - IPv6 requires extensive bare-metal
  - Many legacy applications are not designed to run on IPv6
  - ~70 percent of the end-user devices only support IPv4

MUD

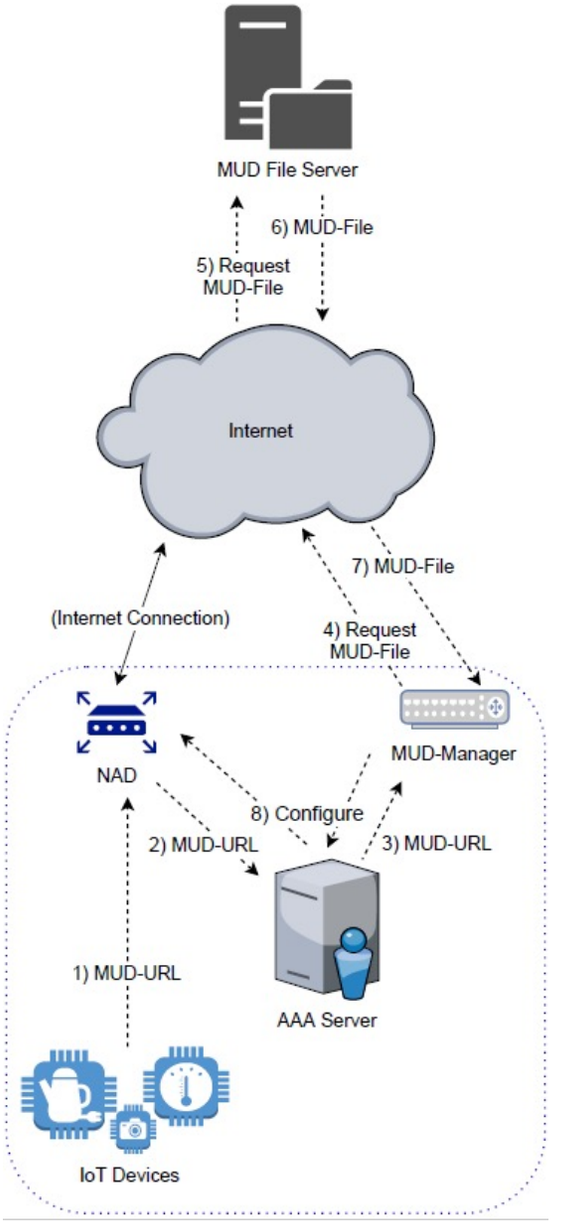
- Recent IETF standard
- Automatically configure devices' access control
- Isolation-based defense

Workflow


MUD File Server
6) MUD-File
5) Request MUD-File
Internet
7) MUD-File
(Internet Connection)
4) Request MUD-File
NAD
MUD-Manager
8) Configure
2) MUD-URL
3) MUD-URL
1) MUD-URL
AAA Server
IoT Devices

MUD-File

- One of the main components of MUD
- May contain hundreds of ACEs (JSON)
- Usability/Acceptability? Difficult to:
  - Read
  - Validate
  - Analyze (interactions)



MUD-File


```
{
  "ietf-mud:mud": {
    "mud-version": 1,
    "mud-url": "https://example.org/tester",
    "last-update": "2019-08-05T20:24:54+00:00",
    "cache-validity": 48,
    "is-supported": true,
    "systeminfo": "This is just an example ",
    "mfg-name": "Example LLC.",
    "documentation": "https://example.org/docs",
    "model-name": "tester",
    "from-device-policy": {
      "access-lists": {
        "access-list": [
          {
            "name": "mud-64733-v4fr"
          }
        ]
      }
    },
    "to-device-policy": {
      "access-lists": {
        "access-list": [
          {
            "name": "mud-64733-v4to"
          }
        ]
      }
    }
  },
  "ietf-access-control-list:acls": {
    "acl": [
      {
        "name": "mud-64733-v4to",
        "type": "ipv4-acl-type",
        "aces": {
          "ace": [
            {
              "name": "cl0-todev",
              "matches": {
                "ipv4": {
                  "ietf-acldns:src-dnsname": "www.example.org",
                  "protocol": 6
                },
              },
              "actions": {
                "forwarding": "accept"
              }
            }
          ]
        }
      },
      {
        "name": "mud-64733-v4fr",
        "type": "ipv4-acl-type",
        "aces": {
          "ace": [
            {
              "name": "cl0-frdev",
              "matches": {
                "ipv4": {
                  "ietf-acldns:dst-dnsname": "www.example.org",
                  "protocol": 6
                },
              },
              "actions": {
                "forwarding": "accept"
              }
            }
          ]
        }
      }
    ]
  }
}
```


{ "ietf-mud:mud": { "mud-version": 1, "mud-url": "https://example.org/tester", "last-update": "2019-08-05T20:24:54+00:00", "cache-validity": 48, "is-supported": true, "systeminfo": "This is just an example ", "mfg-name": "Example LLC.", "documentation": "https://example.org/docs", "model-name": "tester",

"from-device-policy": { "access-lists": { "access-list": [{ "name": "mud-64733-v4fr" }] } },

"to-device-policy": { "access-lists": { "access-list": [{ "name": "mud-64733-v4to" }] } } },

"ietf-access-control-list:acls": { "acl": [{ "name": "mud-64733-v4to", "type": "ipv4-acl-type", "aces": { "ace": [{ "name": "cl0-todev", "matches": { "ipv4": { "ietf-acldns:src-dnsname": "www.example.org", "protocol": 6 }, }, "actions": { "forwarding": "accept" } }] } }, { "name": "mud-64733-v4fr", "type": "ipv4-acl-type", "aces": { "ace": [{ "name": "cl0-frdev", "matches": { "ipv4": { "ietf-acldns:dst-dnsname": "www.example.org", "protocol": 6 }, }, "actions": { "forwarding": "accept" } }] } }] } }

How to make MUD more Usable/Acceptable?

MUD-Visualizer

Goals:

- Protocol Checking to detect errors in MUD-Files
- Optimization of MUD-Files, e.g., overlapping rules
- Visualization of the behavior of the IoT devices and their interactions

Does MUD-Visualizer make MUD more Usable/Acceptable?

Let’s see…

Experiment

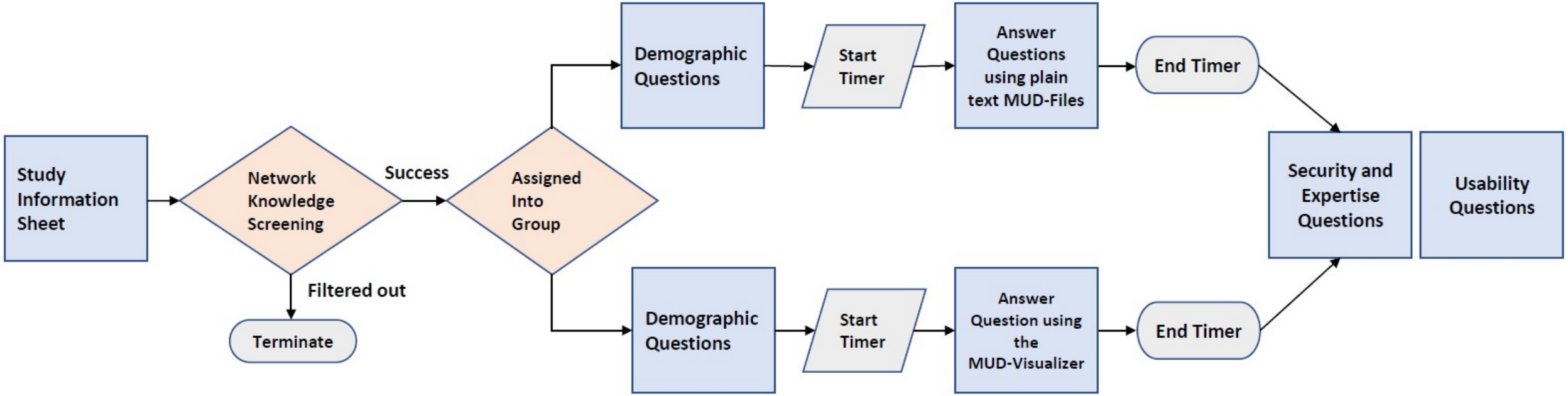
Research Questions on Acceptability

- To what extent does MUD-Visualizer improve the **usability** of the analysis of the MUD-Files?
- How much does MUD-Visualizer affect the **accuracy** of the analysis of the MUD-Files?
- How much does MUD-Visualizer affect the **time** of the analysis of the MUD-Files?
- To what extent does **knowledge of security** affect the accuracy of the analysis of the MUD-Files?

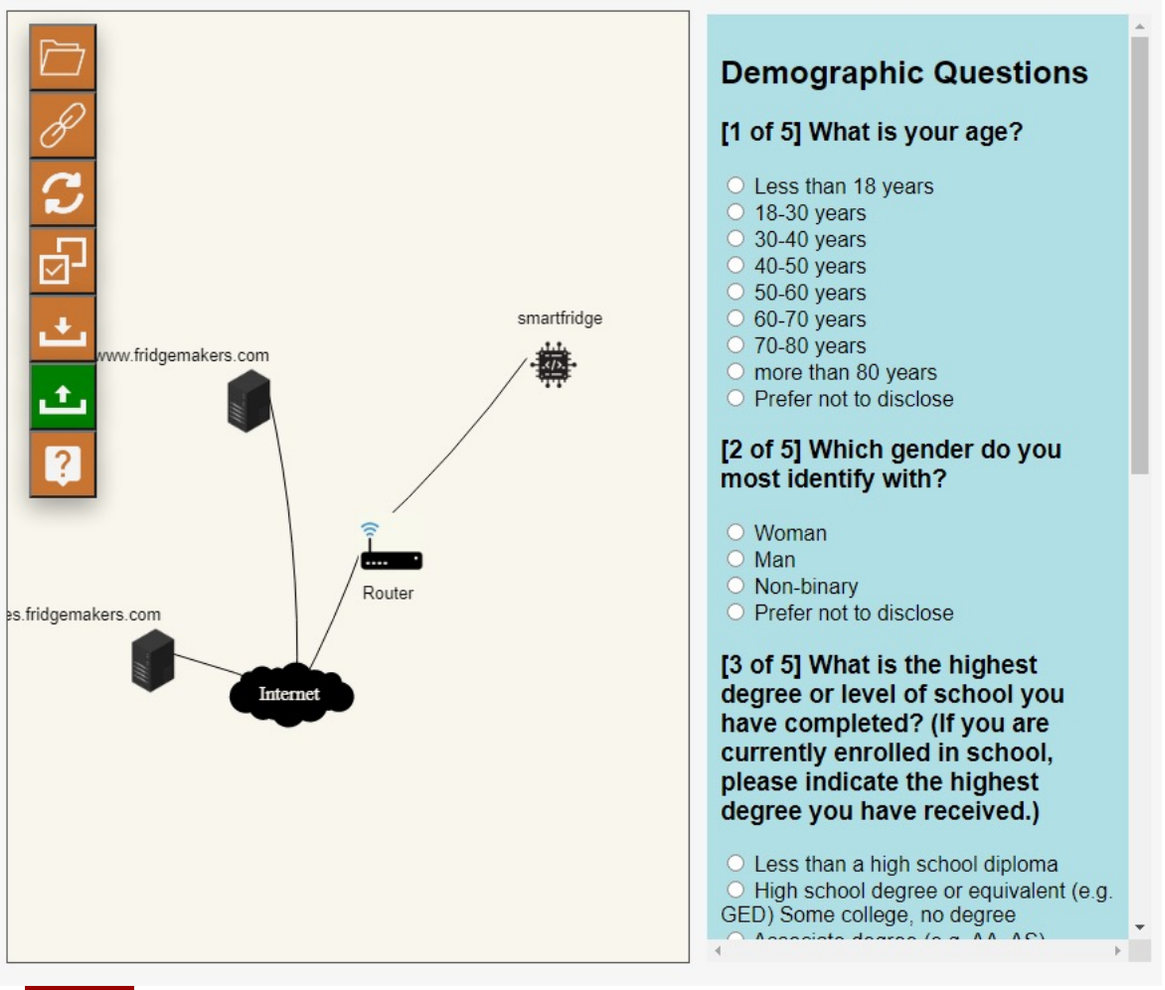
Pilot Study

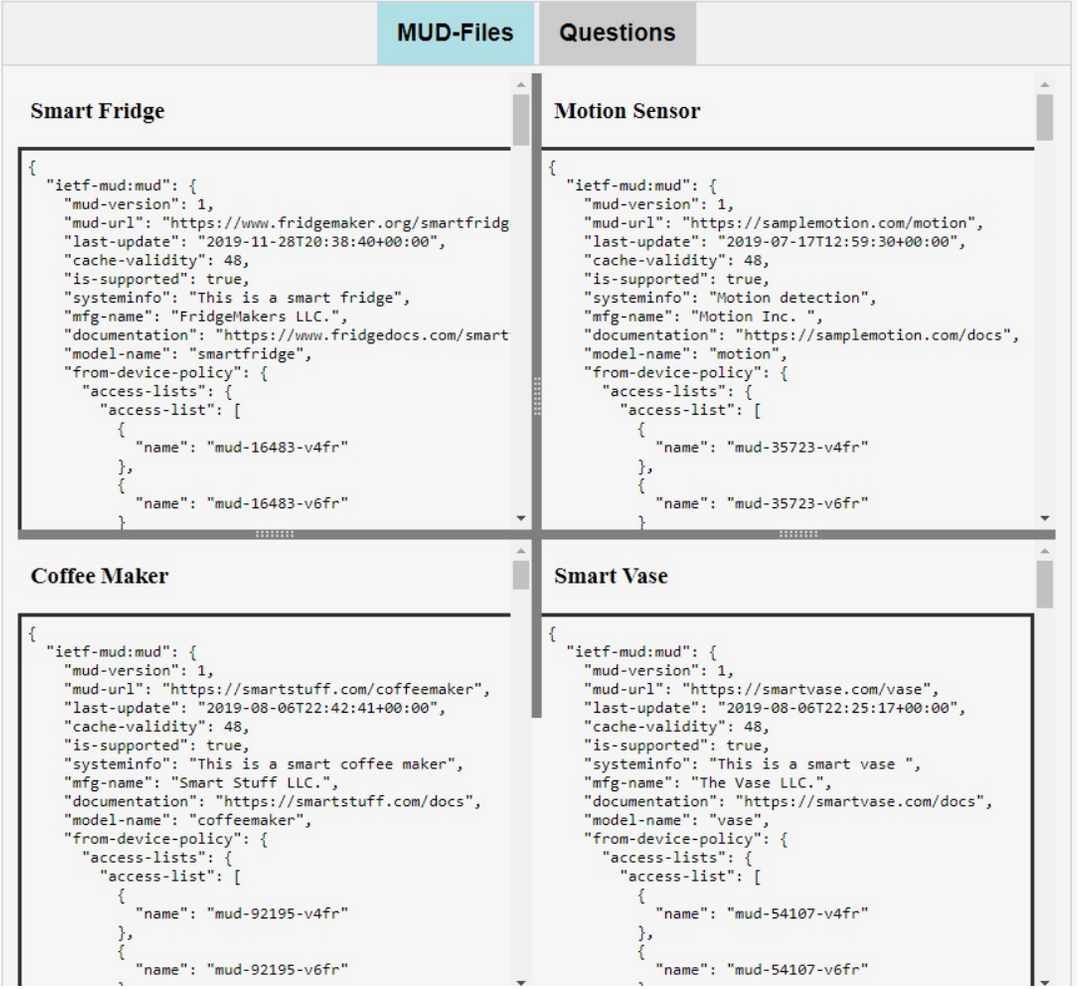
- 8 people in our lab (4 people in each group)
- No screening
- Random assignment to groups
- Confirming study design and adding screening survey

Experiment flow



Experiment Platform





Screening

- To ensure that the participants have the required knowledge of networking
- Was achieved through asking them to parse a partial MUD-File
- The experiment was advertised only to graduate CS students and students in advanced computer networking course

Variables in the Dataset

- [5 Qs] The **Demographic** questions was about age, gender, education, employment status and income [1]
- The **main experiment** questions about analysis of the MUD-Files in two categories:
  - [10 Qs] Number/identity of the nodes that devices allow-listed
  - [13 Qs] Traffic details of the allowed communication in transport and network layer

[1] Henrich, J., Heine, S.J., Norenzayan, A.: Most people are not WEIRD. Nature, 466(7302), 29–29 (2010)

Variables in the Dataset

- Comprised 50 questions in two categories:
  - [40 Qs] A set of computer expertise questions [2]
  - [10 Qs] Usability questions from System Usability Scale (SUS) [3]

[2] Rajivan, P., Moriano, P., Kelley, T., Camp, L.J.: Factors in an End User Security Expertise Instrument. Information & Computer Security (2017)
[3] Brooke, J.: SUS: A “Quick and Dirty” Usability. CRC Press (1996)

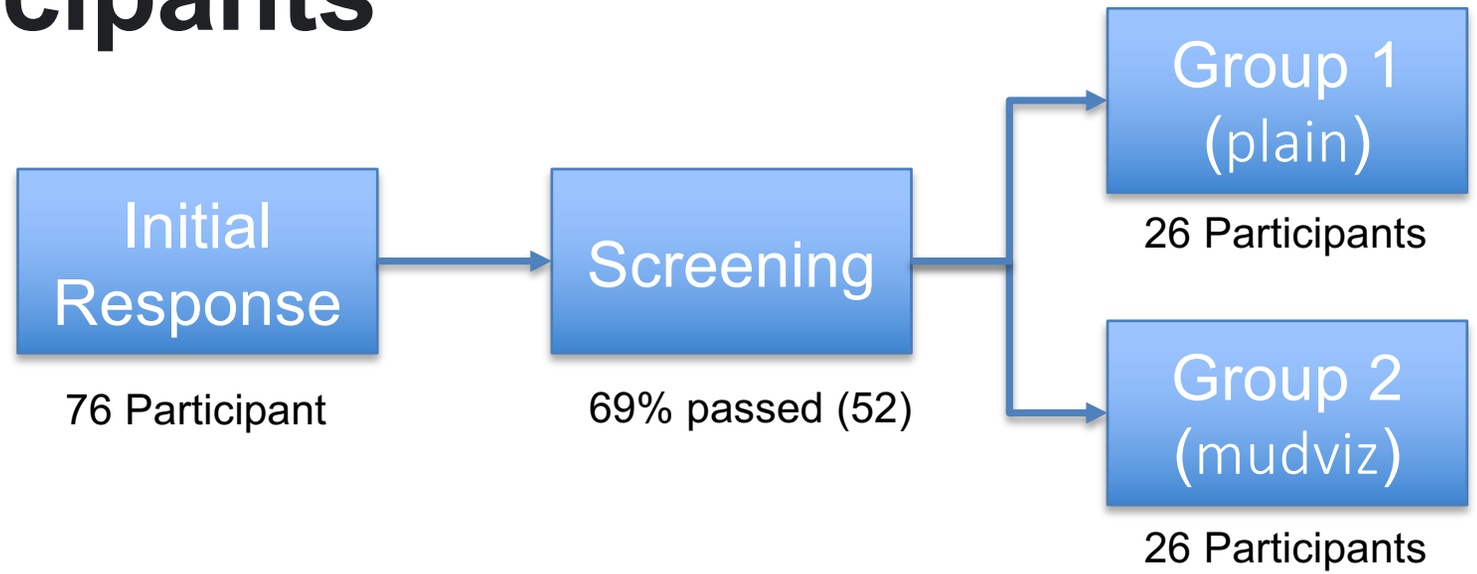
Building the Dataset


```
{
    "I_think_that_I_would_like_to_perform_this_analysis_frequently": "2",
    "I_found_the_analysis_unnecessarily_complex": "10_Strongly_agree",
    "I_thought_the_analysis_was_easy": "10_Strongly_agree",
    "I_think_that_I_would_need_the_support_of_a_technical_person_to_be_able_to_perform_the_analysis":
"1_Strongly_disagree",
    "I_found_the_various_components_in_this_analysis_were_well_integrated": "10_Strongly_agree",
    "I_thought_there_was_too_much_inconsistency_in_this_analysis": "1_Strongly_disagree",
    "I_would_imagine_that_most_people_would_learn_to_use_this_analysis_very_quickly":
"10_Strongly_agree",
    "I_found_this_analysis_very_cumbersome_to_perform": "1_Strongly_disagree",
    "I_felt_very_confident_performing_the_analysis": "8",
    "I_needed_to_learn_a_lot_of_things_before_I_could_get_going_with_the_analysis":
"1_Strongly_disagree",
    "ResponseTime": "107849",
    "workerId":
"aa8a6eb6b149713c33525e6dd0299e8af13c521690b2636fab05e74cc1645cad0357fb0c0345f216f8514d9f9814a5e2993d2
7f65e6da4cc1cbed3e4a8517711",
    "assignmentId": null,
    "valid_participant": null
}
```


- Responses were in JSON
- They were converted to CSV
- They were merged into a single CSV per participant
- We added extra variable indicating the group
- All CSVs were merged to create our dataset

Metrics and Analytical Techniques

Participants



- 41 / 52 were < 30 years old
- > 70% student
- > 96% Bachelor's degree

Perceived Usability

To what extent does MUD-Visualizer improve the **usability** of the analysis of the MUD-Files?

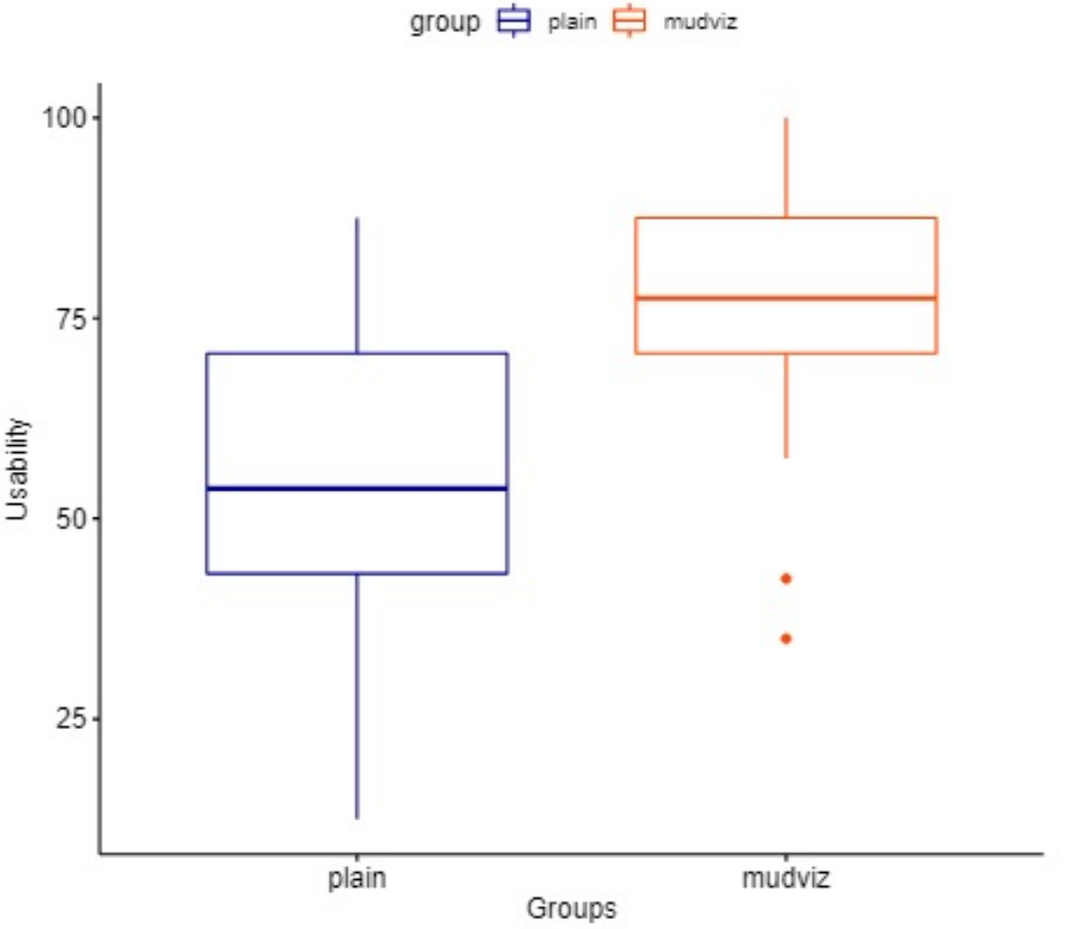
- We used System Usability Scale (SUS) to generate a single usability score out of 100
- An aggregate score of 68 is considered to be average [4]
- We used Shapiro test and determined we cannot assume normality

[4] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An Empirical Evalu-ation of the System Usability Scale.International Journal of Human–ComputerInteraction24, 6 (2008), 574–594

Perceived Usability

To what extent does MUD-Visualizer improve the **usability** of the analysis of the MUD-Files?

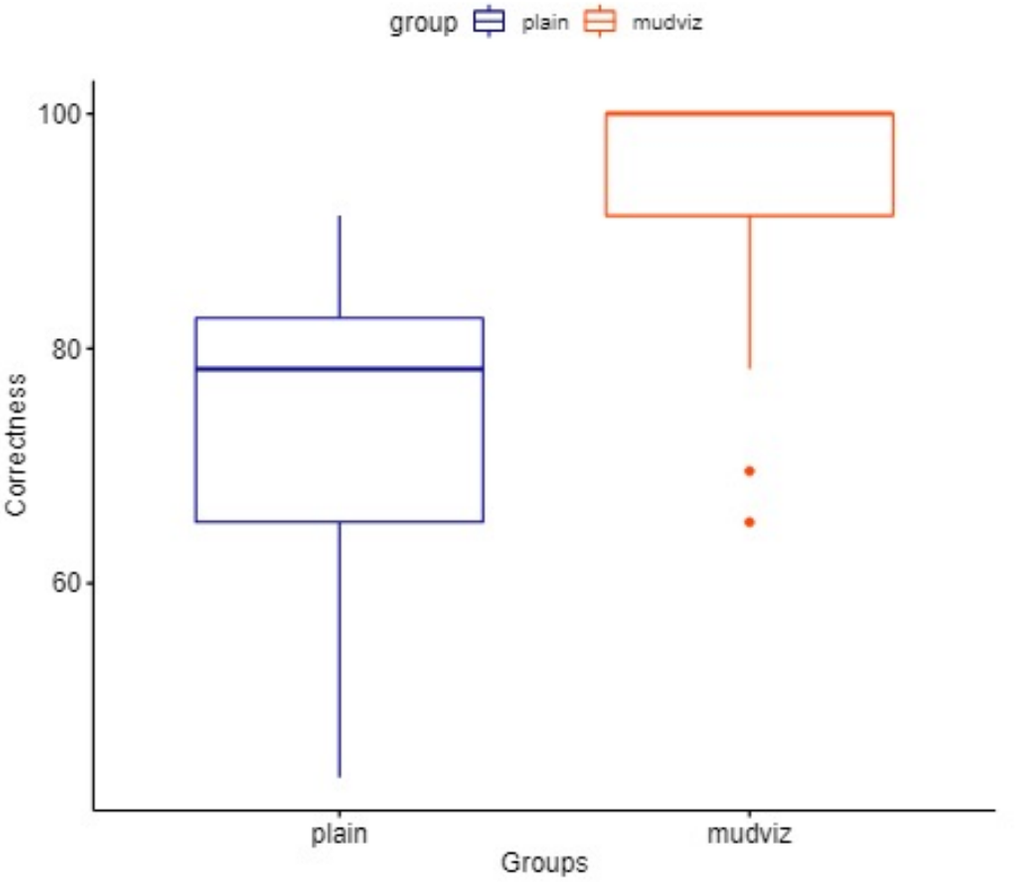
- A non-parametric Mann-Whitney rank-sum test indicated that the usability of MUD-Visualizer was significantly higher than plain text analysis (P-Value = 1.687e-04)



Measured Usability: Total Accuracy

How much does MUD-Visualizer affect the **accuracy** of the analysis of the MUD-Files?

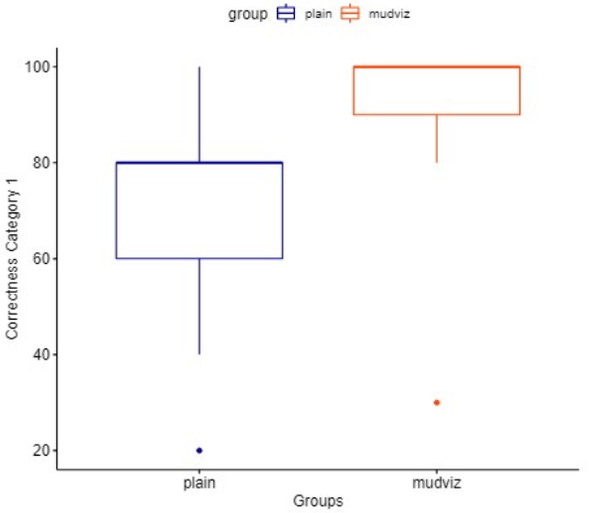
- The different of total accuracy in both groups was also statistically significant (P-Value: 8.70e-05)



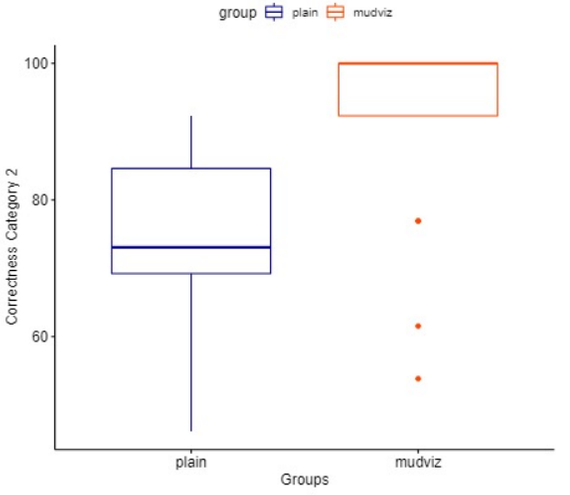
Measured Usability: Accuracy per Group

How much does MUD-Visualizer affect the **accuracy** of the analysis of the MUD-Files?

- We had two groups:
  - Nodes and communications
  - Traffic Details
- Wilcoxon Rank-Sum Test showed the distance is statistically significant: (P-Values: 4.203e-04 and 4.268e-04)



(a) Correctness for Nodes & Communications



(b) Correctness for Traffic Details

Measured Usability: Effect Size

How much does MUD-Visualizer affect the **accuracy** of the analysis of the MUD-Files?

- We calculated the effect size using Cohen's D formula
- As a rule of thumb, the effect size between 0.5 and 0.8 is considered large [5]

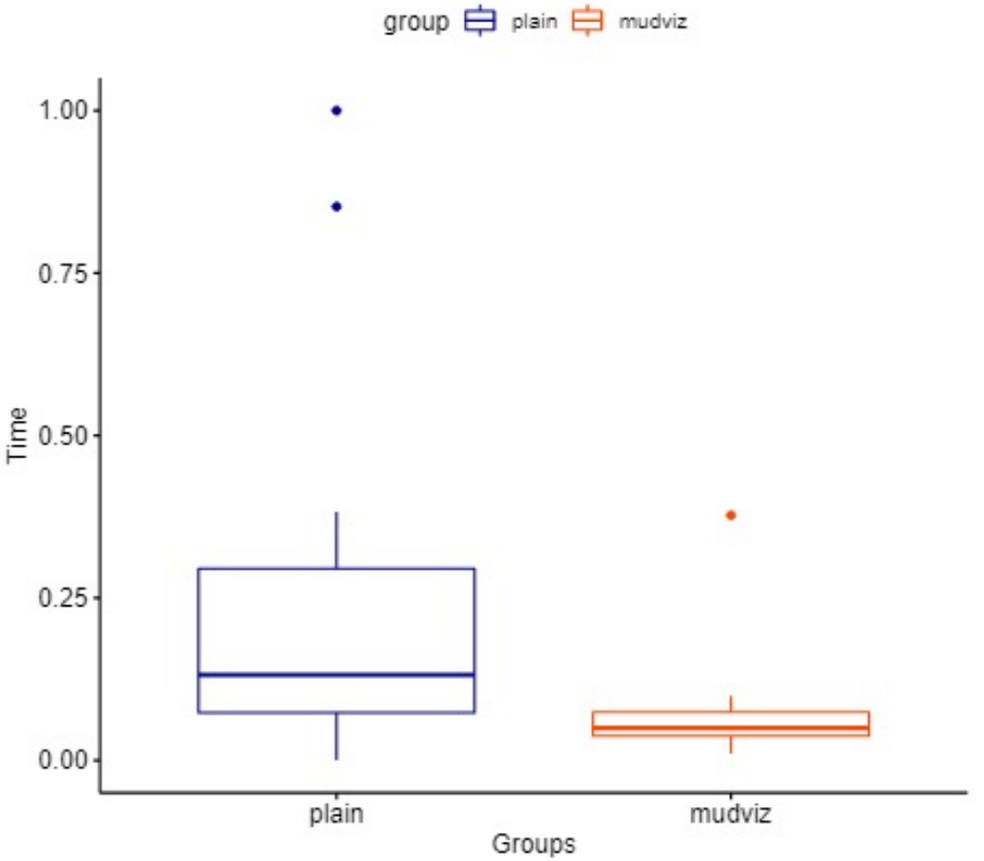
Variables Compared	Odds Ratio	Effect Size	P-Value
Comparison of overall accuracy between the two groups	1.3	0.77	8.70e-05
Comparison of accuracy for `Nodes & Communications`	1.2	0.69	4.20e-04
Comparison of accuracy for `Traffic Details`	1.4	0.81	3.59e-05
Comparison of time to task completion between the two groups	1.2	0.69	4.12e-04

[5] Jacob Cohen. 2013.Statistical Power Analysis for the Behavioral Sciences. Academicpress.

Measured Usability: Time

How much does MUD-Visualizer affect the **time** of the analysis of the MUD-Files?

- Wilcoxon Rank-Sum Test showed that this difference is statistically significant
- Time to task completion also had a large effect size of 0.69



[5] Jacob Cohen. 2013.Statistical Power Analysis for the Behavioral Sciences. Academicpress.

Measured Usability: Effect of the knowledge of Security

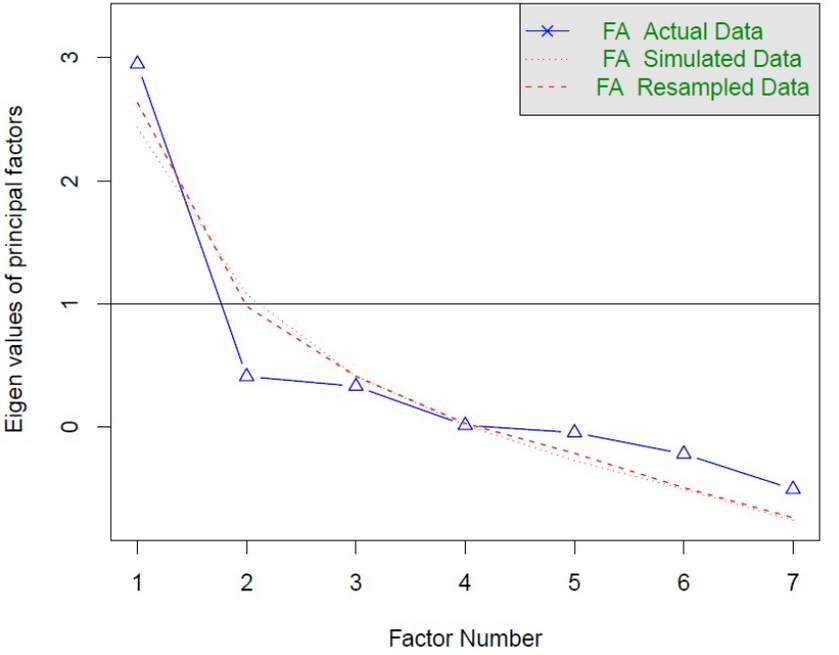
To what extent does **knowledge of security** affect the accuracy of the analysis of the MUD-Files?

- We measured knowledge based on the answer of participants to questions about:

Phishing, Certificates, SQL commands, Intrusion Detection Systems, Port 80, Website markers for security, Defining IoT, Access Control

- The factor TotalKnowledge was a combination of four factors:

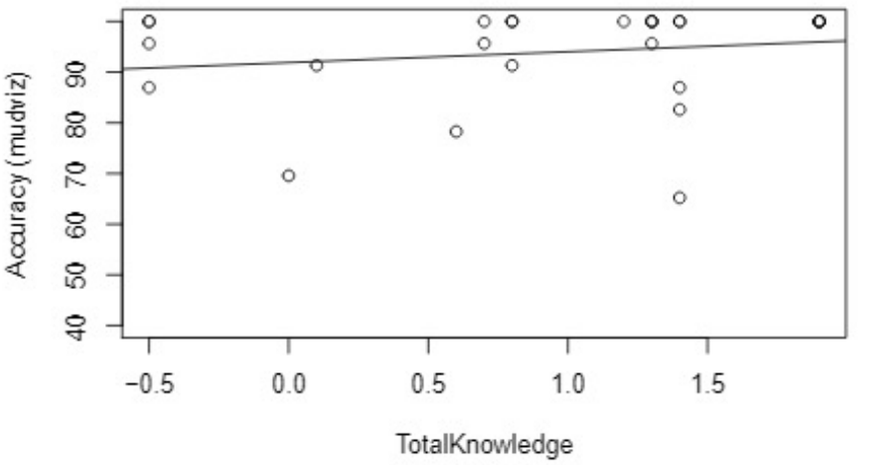
$$\text{TotalKnowledge} \leftarrow (-0.5 * \text{cert}) + (0.6 * \text{sql}) + (0.6 * \text{ids}) + (0.7 * \text{p80})$$



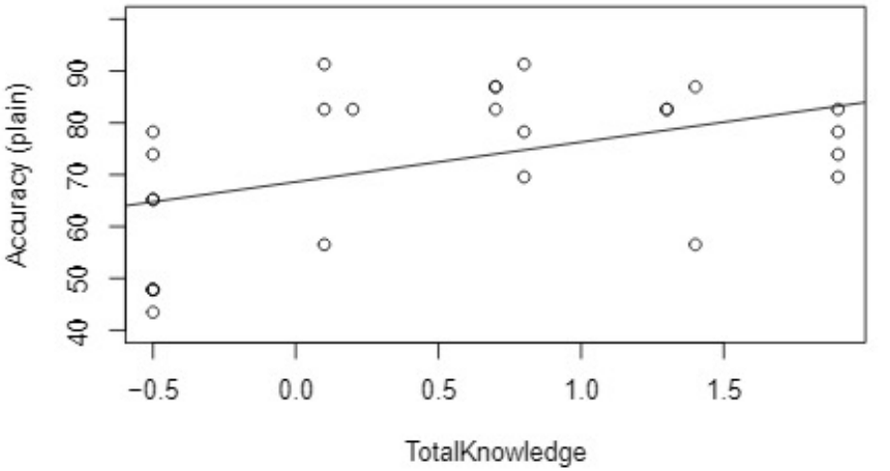
Measured Usability: Effect of the knowledge of Security

To what extent does **knowledge of security** affect the accuracy of the analysis of the MUD-Files?

- The effect of the **knowledge** on accuracy was measured by performing a linear regression
- The effect of security knowledge is significant in the plain group (P-Value 0.0164) but not in the mudviz group (P-Value 0.406)



(a) Accuracy v. TotalKnowledge (mudviz)



(b) Accuracy vs TotalKnowledge (plain)

Discussion

Questions

- How important is producing intermediate results?
  - Prescreening questions seemed necessary in our case, what about in other cases, e.g., open-source community?
- What did you try that did not succeed?
  - We first used embedded Google docs for survey responses, but later on we noticed we cannot measure time with it (easily)
  - We used non-parametric Student t-test, but after using Shapiro test and confirming non-normal distribution, we used Wilcoxon Rank-Sum test

Questions

- What can be learned from your methodology?
  - Can we use the same methodology for other tools?
  - How much would qualitative questions add value in case of security tools?
  - How would you choose your participants?

Takeaways

- Usability and acceptability should be considered as one of the main components during the standard design
- In case of MUD, we developed MUD-Visualizer and conducted a survey to measure the efficacy of MUD-Visualizer
- The below average SUS score of the plaintext MUD-Files was an indication of the challenges in the usability
- With MUD-Visualizer the analysis of MUD-Files can be done with higher accuracy in a shorter amount of time
- Also, when MUD-Visualizer is not used, deeper security knowledge is required to read and analyze the MUD-Files accurately
- We are considering mixed methods (qualitative and quantitative) for our next work (e.g., for SBoM)

Takeaways

- Usability and acceptability should be considered as one of the main components during the standard design
- In case of MUD, we developed MUD-Visualizer and conducted a survey to measure the efficacy of MUD-Visualizer
- The below average SUS score of the plaintext MUD-Files was an indication of the challenges in the usability
- With MUD-Visualizer the analysis of MUD-Files can be done with higher accuracy in a shorter amount of time
- Also, when MUD-Visualizer is not used, deeper security knowledge is required to read and analyze the MUD-Files accurately
- We are considering mixed methods (qualitative and quantitative) for our next work (e.g., for SBoM)

Thank You for Listening!

vafandal@iu.edu
@vafandal