Computer Security: A Signaling Game Approach

Keywhan Chung

Professor Ravishankar Iyer and Professor Zbigniew Kalbarczyk

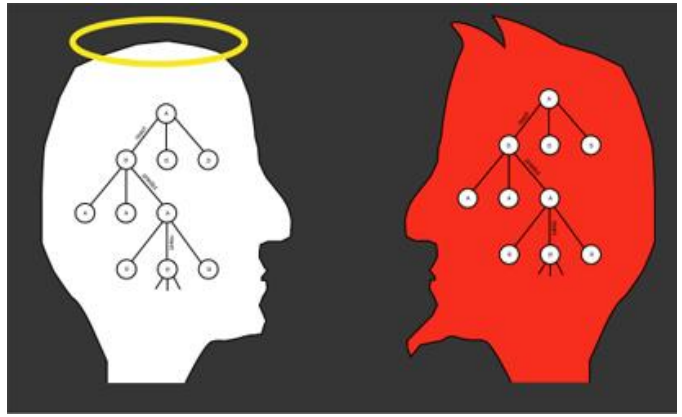
In collaboration with Dr. Charles Kamhoua, Dr. Kevin Kwiat at AFRL

Mar. 16, 2016

Hard to Differentiate Attacks

Signature Based Detection → Variations of Attacks → Anomaly Based Detection

Anomaly Based Detection → Reduce the **significance of abnormality**

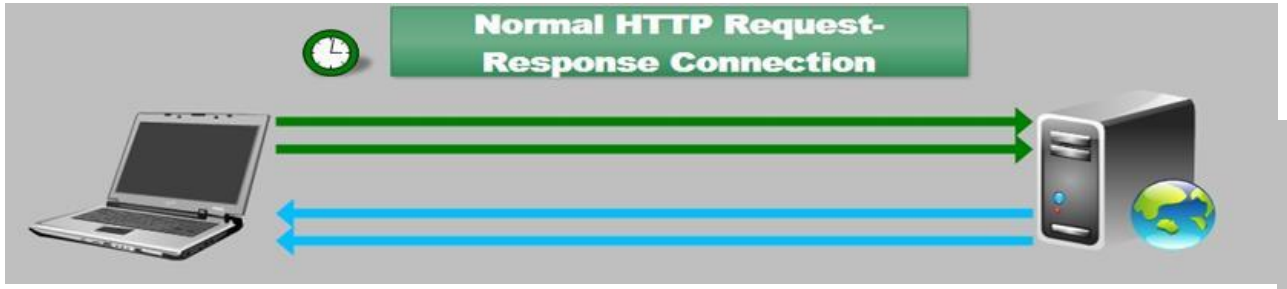


Increased
FALSE NEGATIVE and
FALSE POSITIVE rate

Can a game theoretic **'RATIONAL DECISION MAKING'** contribute?

Example: slowHTTP based DoS

- Traditional DoS on the **NETWORK** layer
- Exploits the **APPLICATION LAYER**
  - ***Silent & Small:*** *No significant trace* to determine malicious intention
  - Less resources required for equivalent impact
- Said that it can **EASILY** be prevented with <u>appropriate the server configuration</u>
- *E.g., slow HTTP GET, slow HTTP POST, slow HTTP header*



incomplete request

Limitations for Security

- Often said that it can **EASILY** be prevented with appropriate the server configuration

However,

- **limited control** over each machine:
  Cannot assure that all applications are hardened.

Need for a method to handle the attack without control over all servers

Previous Work

- Assign captcha challenge to check human behind the request
- Assign resource heavy challenge to client to bring down the client if attempting DoS
- Setup a TIMEOUT (firewall etc. if not on application itself)

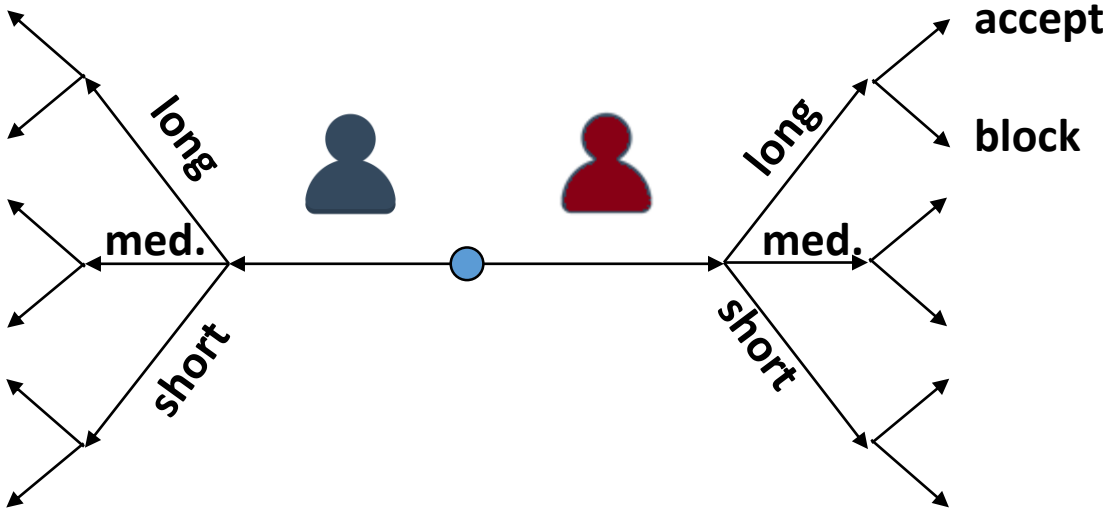
Benign users are affected. Need decision on WHO

Signaling Game

- **Two players w/ conflict of interest**
  - Sender: has types unknown to opponent. Chooses msg to be sent to receiver
  - Receiver: Receives msg from sender and guesses type of sender and chooses action
- **Message**: signal for the receiver to guess the type of sender. Can be a deceit.

In our security game,

- **User**: benign or attacker.
  - Attacker: optimize decision on msg.
  - Benign: no rationality.
- Connection time works as the signal (message)
- **Defender**: monitors connection time and determines response (user type unknown)



Objective

An automated decision model that can protect the system from ‘Hard to Differentiate’ attacks

Real data based game model

Show a practical implementation of a Game Theory based approach for Cyber Security

Preliminary Data Analysis

ts	id.orig_h	id.orig_p	id.resp_h	id.resp_p	duration	orig_bytes	resp_bytes	conn_state	orig
1426395590	141.142.169.3	49662	141.142.2.2	53	0.000224	180	714	SF	US
1426395596	78.186.4.159	35083	141.142.74.100	23	2.976723	0	0	S0	TR
1426395595	107.5.18.180	60253	143.219.110.22	389	-	-	-	S0	US
1426395540	141.142.141.2	44997	149.165.225.1	33447	-	-	-	S0	US

Understanding Benign Traffic

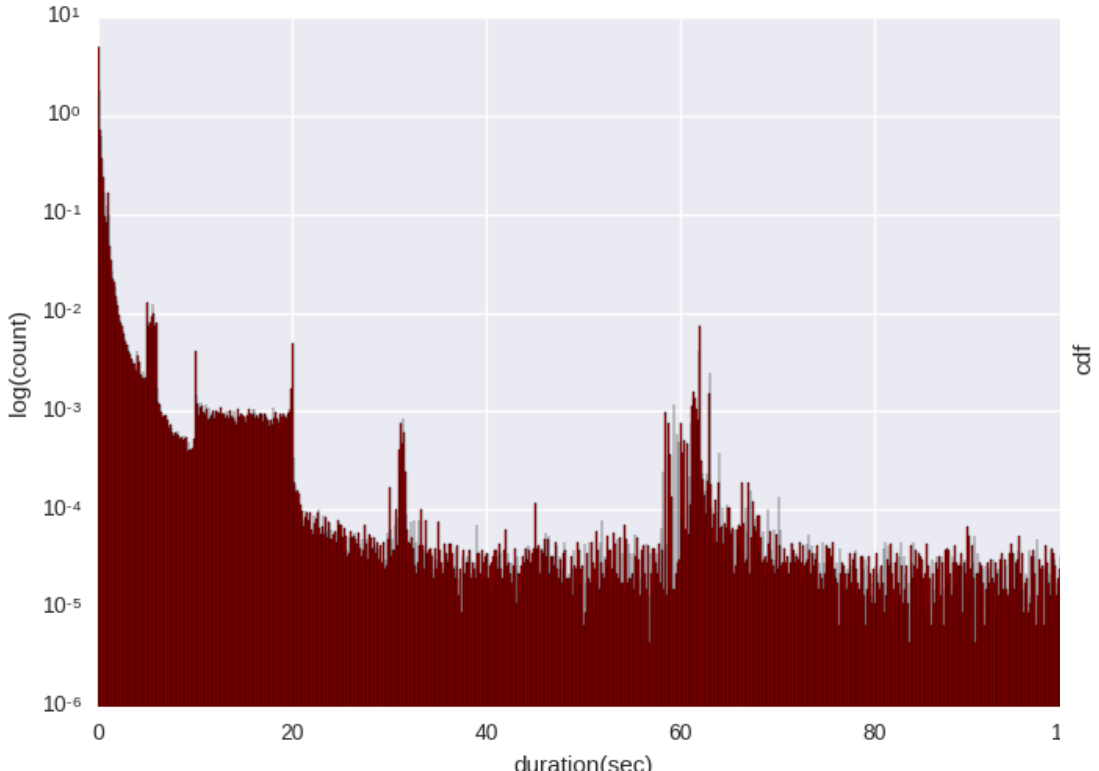
- Web server traffic @NCSA from Jan. 2015 ~ Jun. 2016
  - timestamp, origin / destination (IP and port), protocol, connection duration, connection state, country of origin, size of byte, etc.
- 1M HTTP connections to / from the NCSA website
- No significant sign of NCSA under a slow DoS attack (nothing reported as well)
  - Suspicious Connections: Abnormally long connections with No Bytes actually transferred (0.1%)

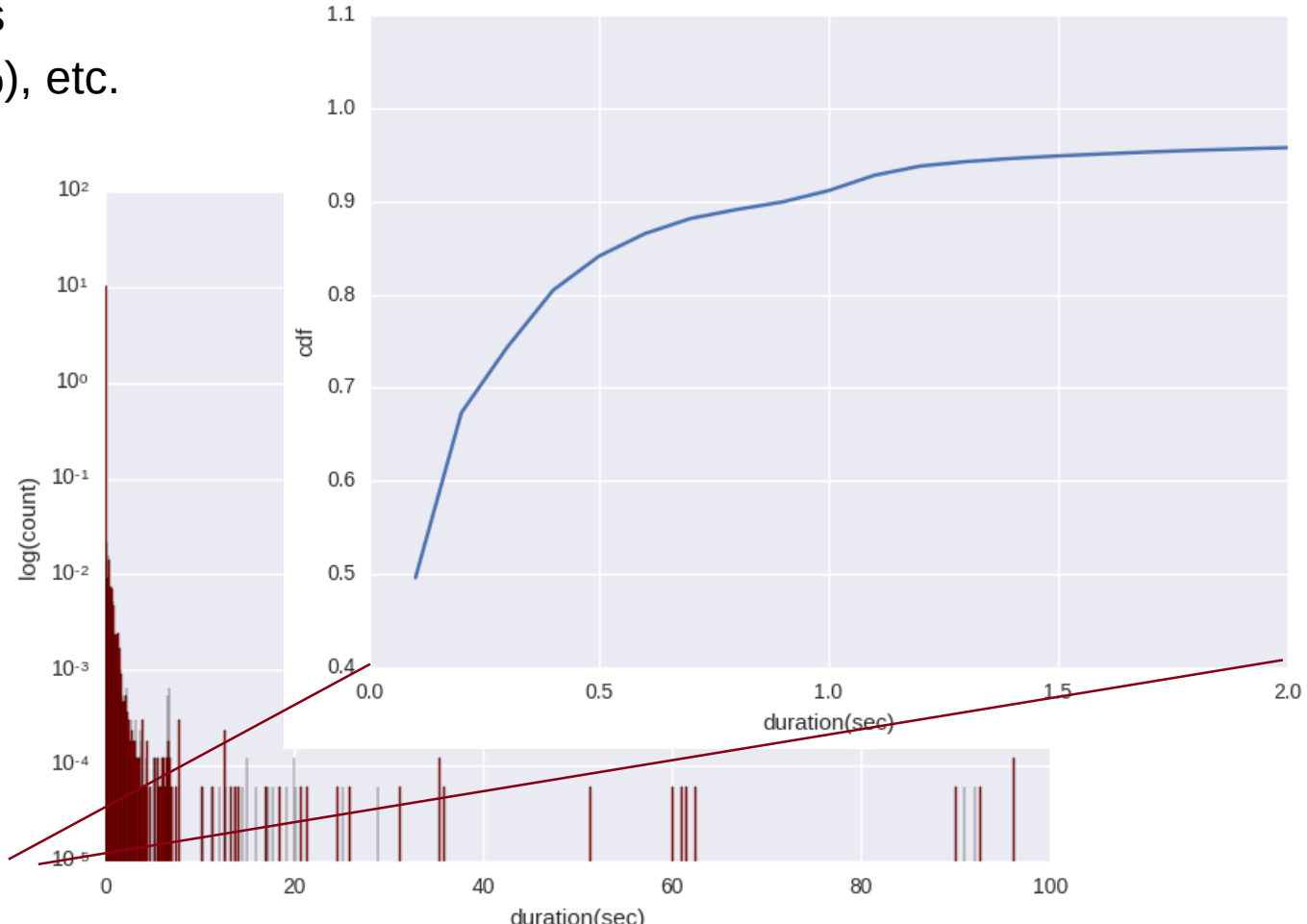
Understanding Attack Traffic

- Run an attack script (slowhttptest) on a victim server
- Monitor the time to failure (of the victim) upon different parameter configurations
- We fix the number of packets (and rate) to assure equivalent visibility

Preliminary Data Analysis

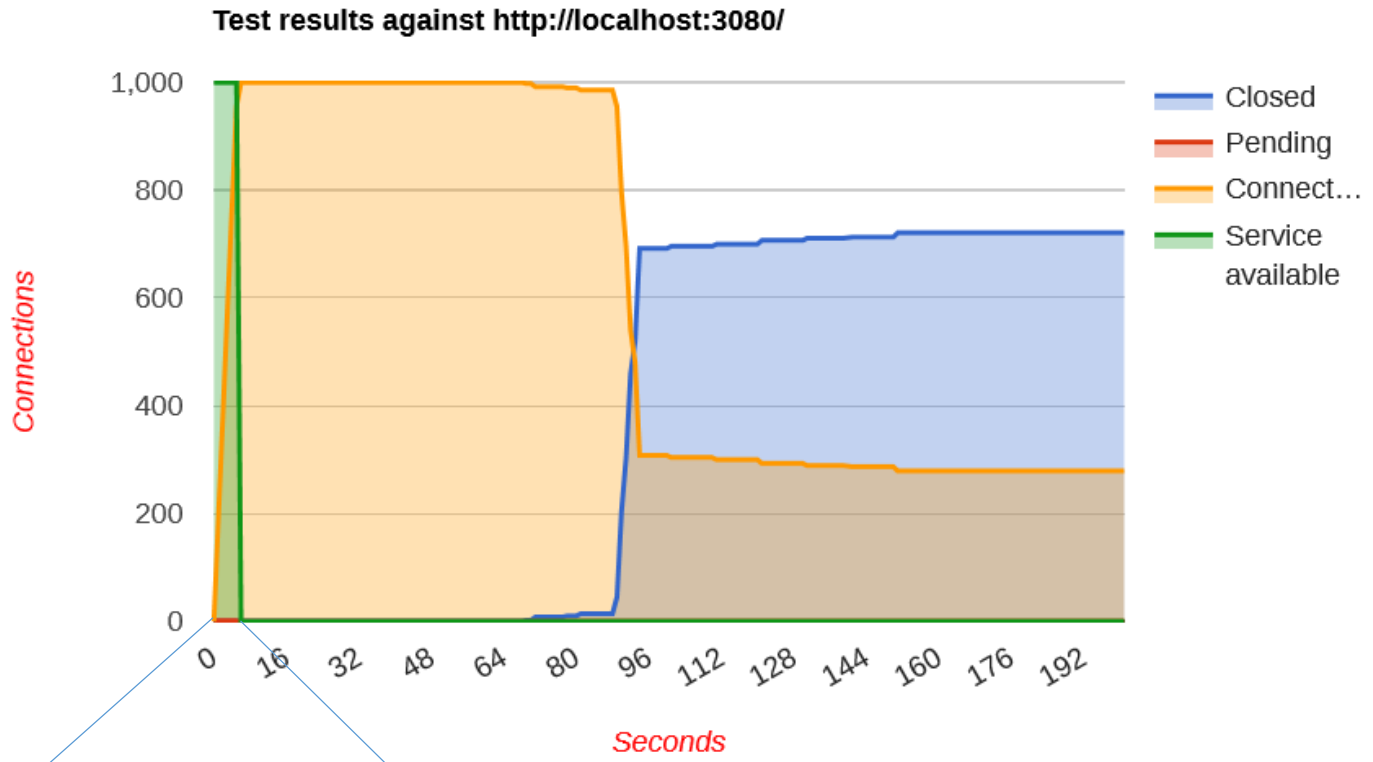
- 212 different originating countries
  - 85% of the traffic is from Top10 countries
  - US (29.81%), GB (27.75%), RU (10.84%), etc.





Preliminary Data Analysis

Test type	SLOW HEADERS
Number of connections	1000
Verb	GET
Content-Length header value	4096
Extra data max length	52
Interval between follow up data	10 seconds
Connections per seconds	200
Timeout for probe connection	3
Target test duration	240 seconds
Using proxy	no proxy

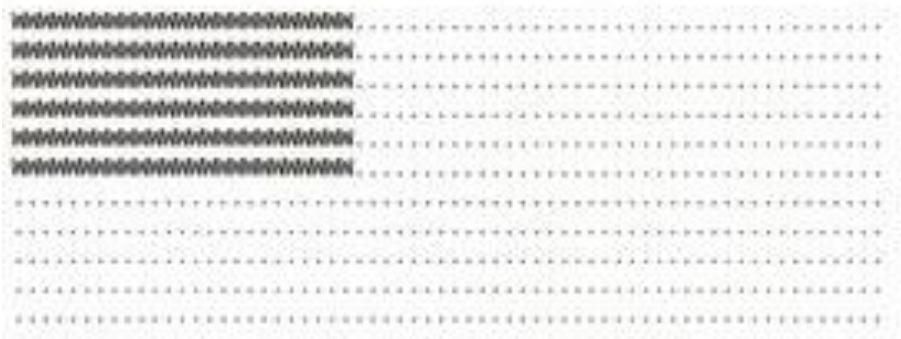


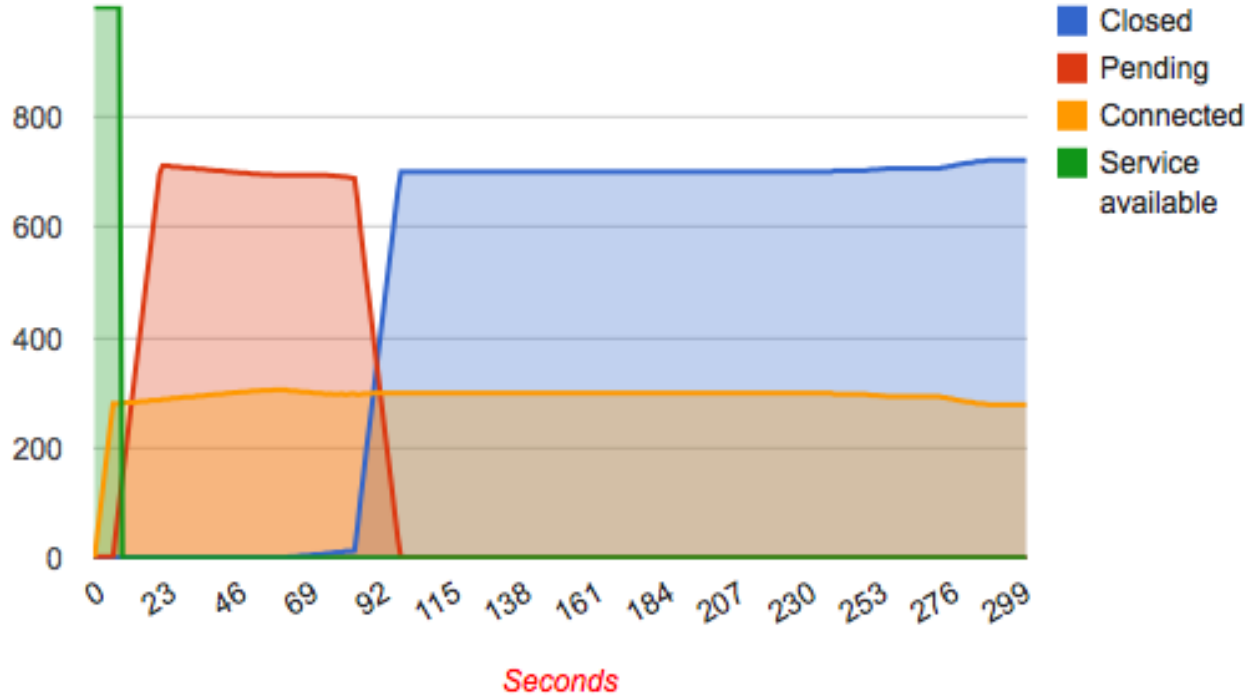
TIME to FAILURE

Preliminary Data Analysis

Test type	SLOW READ
Number of connections	1000
Receive window range	1 - 100
Pipeline factor	10
Read rate from receive buffer	5 bytes / 1 sec
Connections per seconds	50
Timeout for probe connection	5
Target test duration	300 seconds

Server uptime: 1 day 13 hours 29 minutes 54 seconds
Total accesses: 30624 - Total Traffic: 99.2 MB
CPU Usage: u4.12 s.66 cu0 cs0 - .00354% CPU load
.227 requests/sec - 770 B/second - 3396 B/request
150 requests currently being processed, 0 idle workers



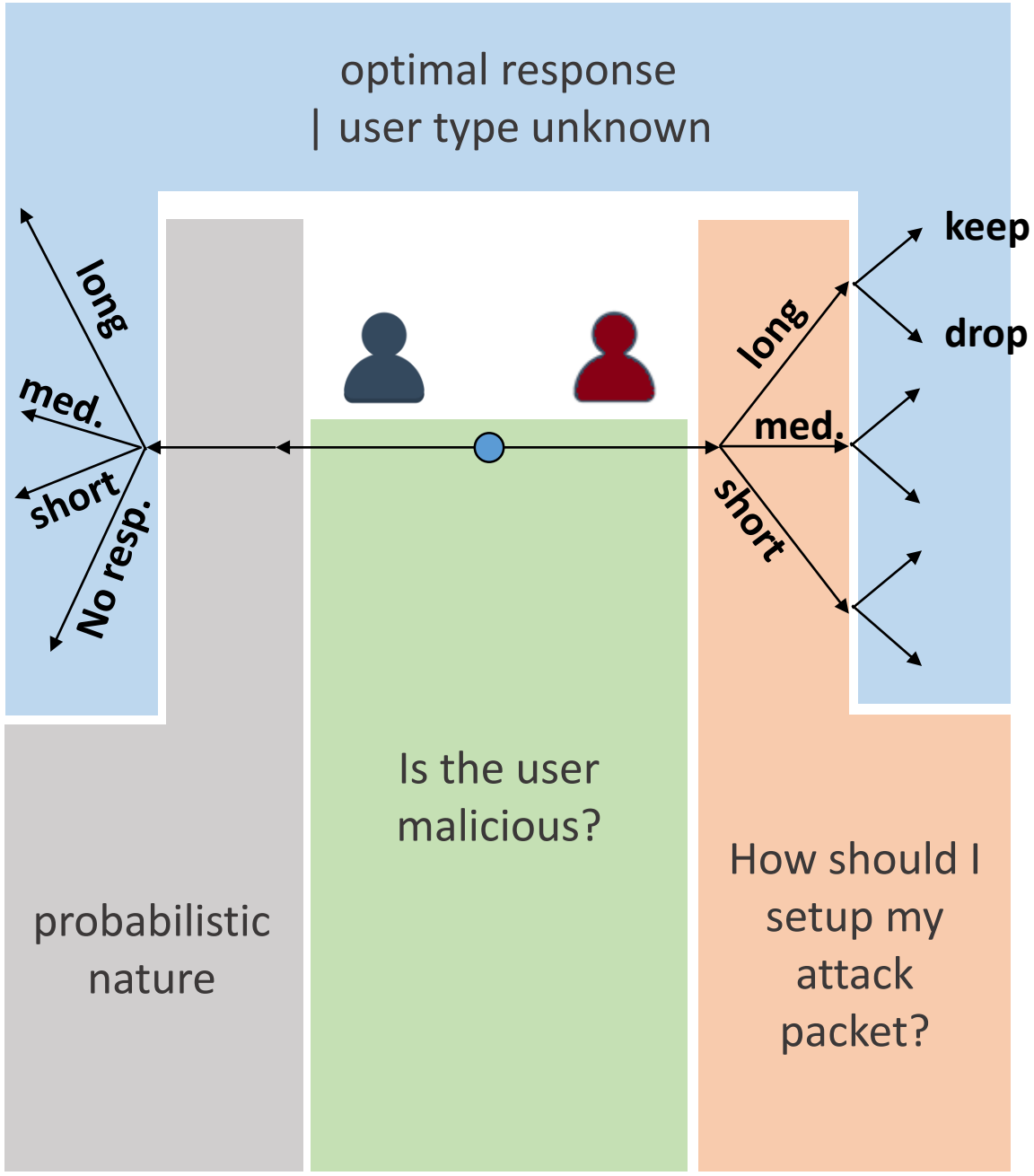


Game Model

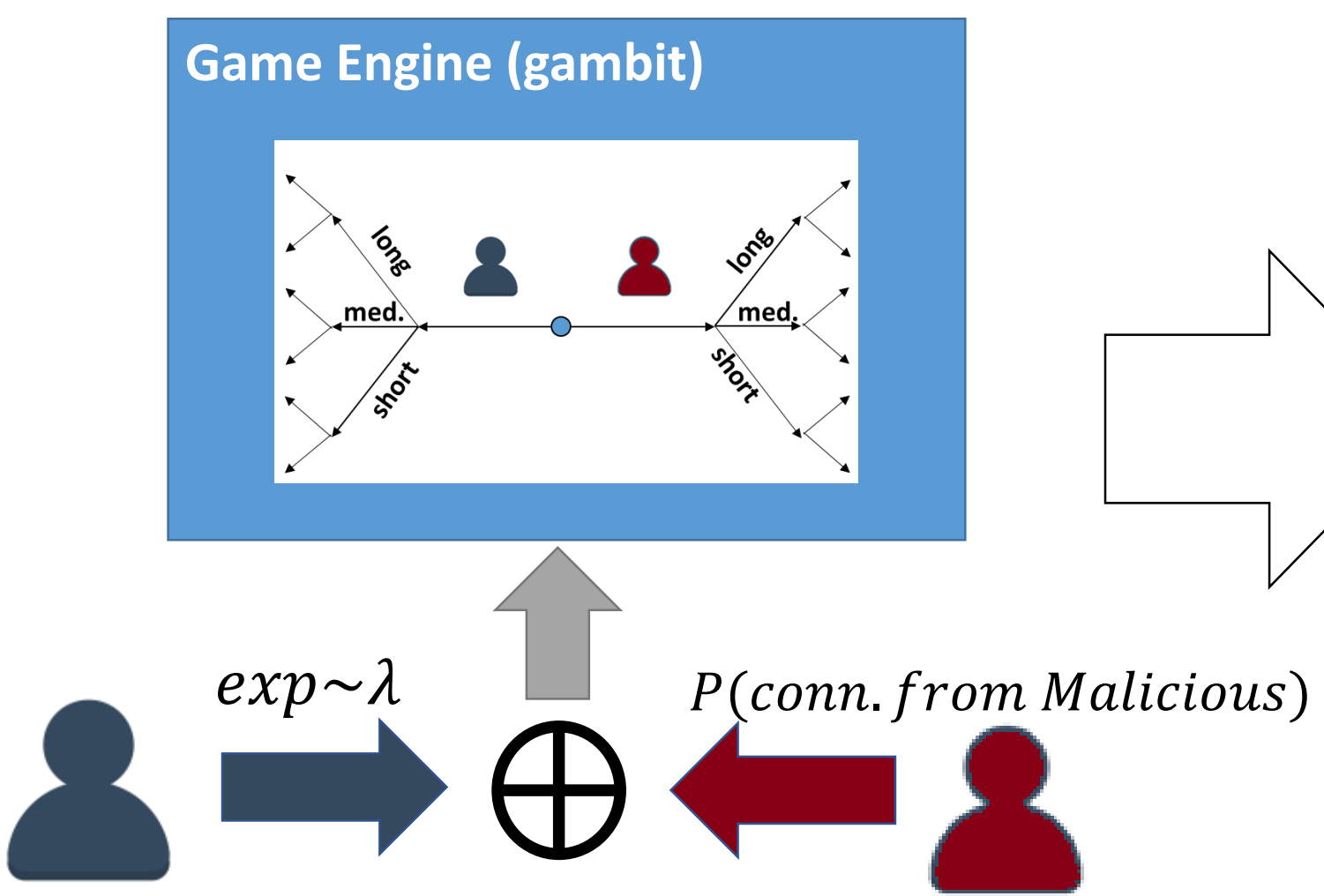
- Uncertainty: **USER TYPE**
- Attacker: optimize connection duration
- Defender: how to handle the packet, given the observation (duration)

- **Reward Model**

	Short	Med Length	Long
TP	±0	±0	±0
FP	± Exp. Users > THR_S	± Exp. Users > THR_M	± Exp. Users > THR_L
TN	±0	±0	±0
FN	± Exp. Users	± Exp. Users	± Exp. Users



Simulation for Accuracy Analysis



Conn #	User	Response
1	Benign	Keep
2	Benign	Keep
3	Benign	Block
4	Benign	Keep
5	Malicious	Block
6	Malicious	Keep
7	Benign	Keep

Simulation Results (1)

- **P(connection from malicious) = 0.01% (suspicious conn. / overall conn)**

Benign (99.99%)		Attacker (0.01%)	
Defender	User	User	Defender
Respond (0.01%)	Long(0.01)	Long(0.05)	Respond (0.01%)
No Respond (99.99%)			No Respond (99.99%)
Respond (0.01%)	Med(0.09)	Med(0.15)	Respond (0.01%)
No Respond (99.99%)			No Respond (99.99%)
Respond (0.01%)	Short(0.9)	Short(0.8)	Respond (0.01%)
No Respond (99.99%)			No Respond (99.99%)

- **High False Negative Rate (most packets are missed)**
- Attacker prefers a short connection (High impact for false positives)
- **No optimal value for P(connection from malicious)**

Simulation Results (2)

- **P(connection from malicious) = 50%**

Benign (50%)		Attacker (50%)	
Defender	User	User	Defender
Respond (33.98%)	Long(0.01)	Long(0.00)	Respond (33.98%)
No Respond (66.02%)			No Respond (66.02%)
Respond (41.77%)	Med(0.09)	Med(0.03)	Respond (41.77%)
No Respond (58.23%)			No Respond (58.23%)
Respond (57.25%)	Short(0.9)	Short(0.97)	Respond (57.25%)
No Respond (43.75%)			No Respond (43.75%)

- **Issues remain for accuracy**
- No optimal value for P(connection from malicious)

Discussion & Future Work

- **P(connection from malicious) has significant impact on the game analysis**
  - **Unknown to the defender from monitoring**
  - **Sensitivity testing for different P values**
  - **Need for a preliminary analysis to derive the optimal P:**
    results from a ML based detector can be a possibility

- **Accuracy of the reward model**
  - **Unlike Q-Learning model, the reward model needs to properly represent the interaction**
  - **Need for a verification on the reward model:**
    factor the impact of difficulty of detection into reward model

- Test on timeliness and accuracy on a real system

Implementation in a Real System

- Q. Is the decision made in a timely manner?
- Q. How accurate is the detection?