# NetCheck: Network Diagnoses from Blackbox Traces

Yanyan Zhuang\*, Eleni Gessiou\*, Fraida Fund\*, Steven Portzer®, Monzur Muhammad\*, Ivan Beschastnikh\*, Justin Cappos\*

(\*) New York University, (^) University of British Columbia, (@) University of Washington

## Goal

- Find bugs in networked applications
  - Large complex unknown applications











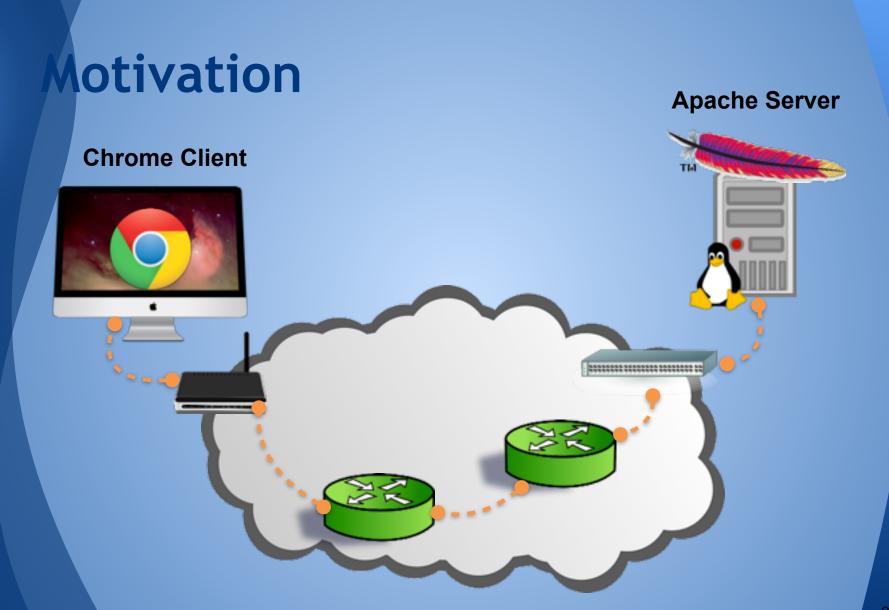
Large complex unknown networks

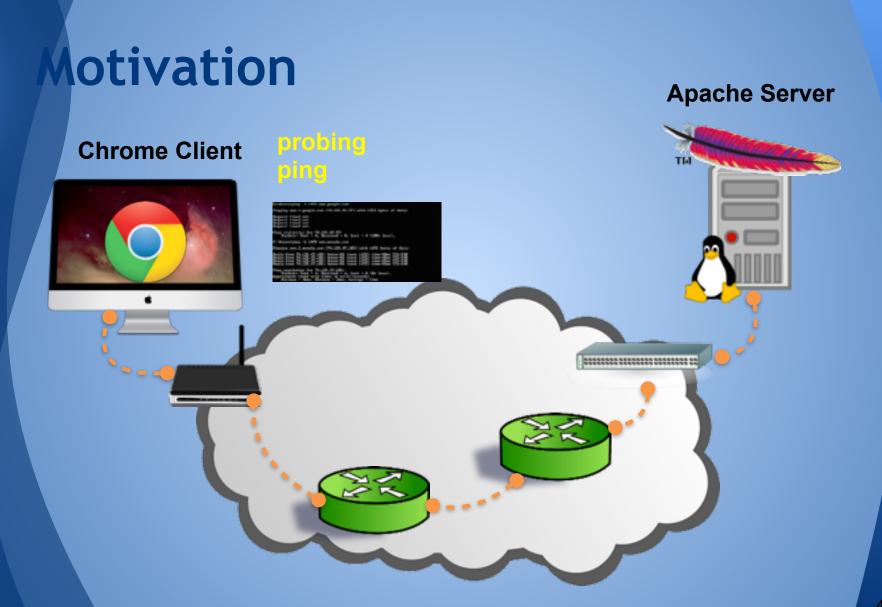




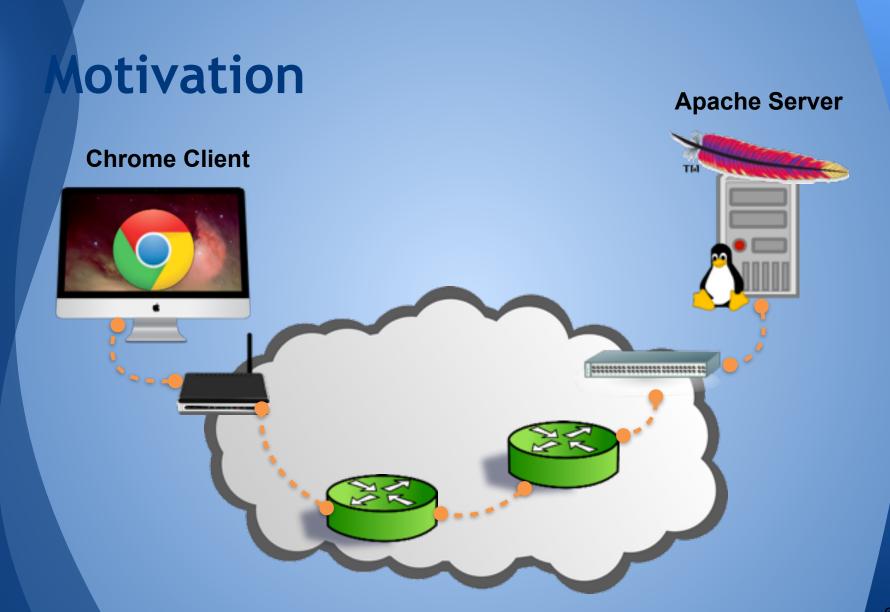


Understandable output / fix



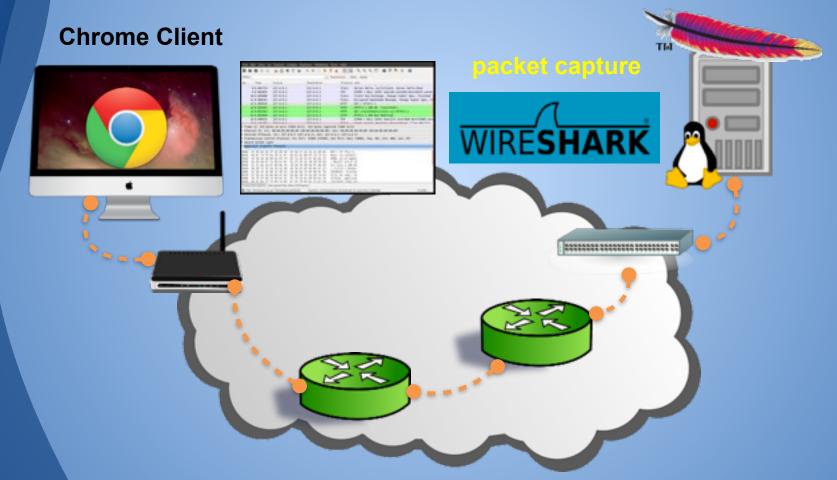


# otivation **Apache Server** Different traffic (ICMP) probing **Chrome Client** Often different result



# Motivation

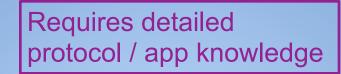
#### **Apache Server**



# Notivation



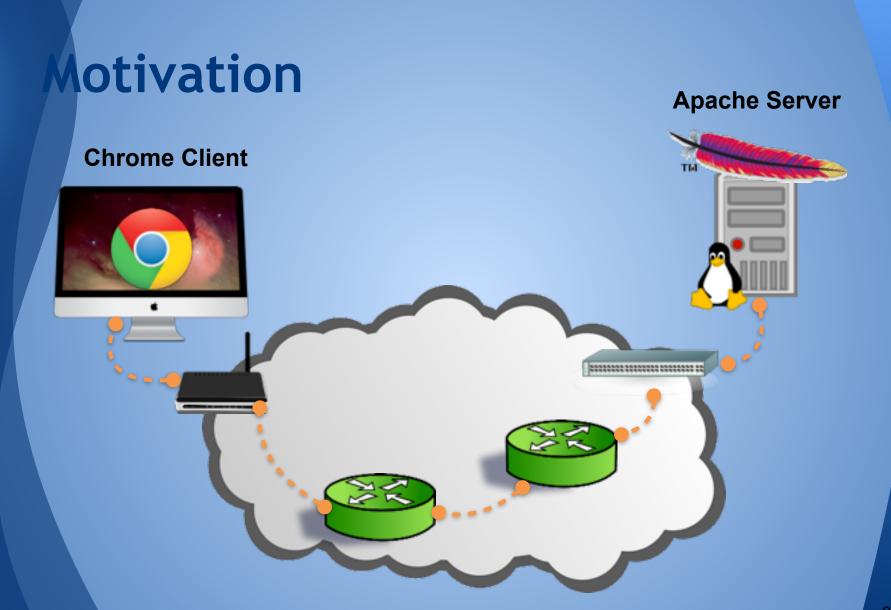


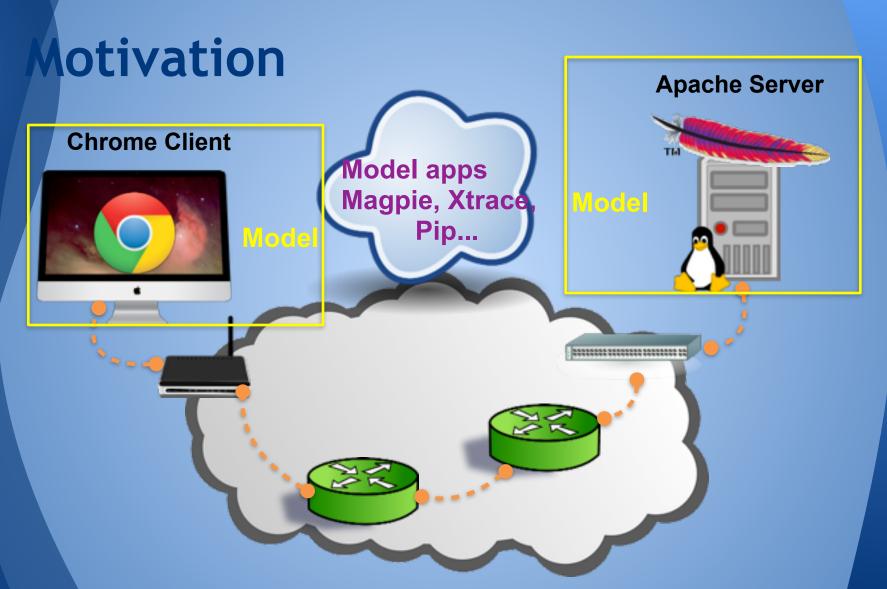


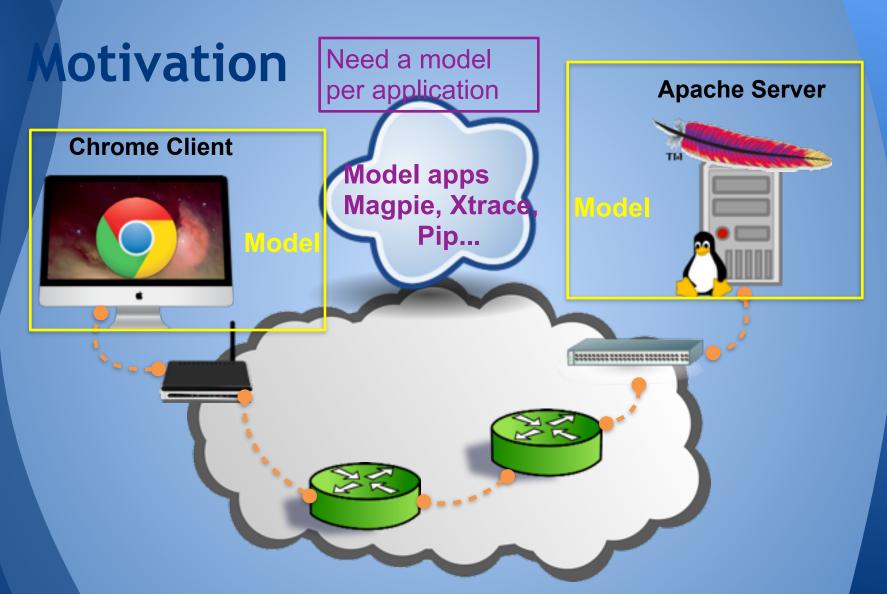


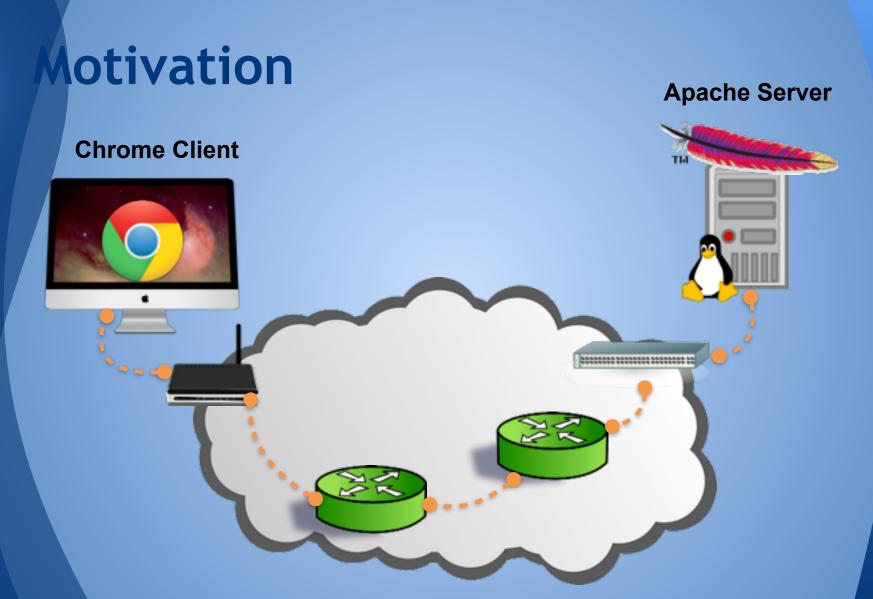
#### **Apache Server**











# Motivation

**Chrome Client** 



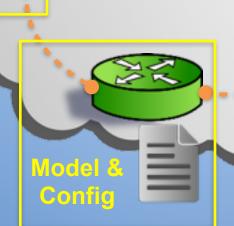
Header Space Analysis, etc.

Network Config Analysis

#### **Apache Server**



Model & Config





........

Model &

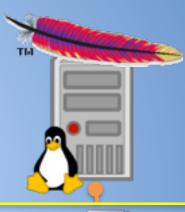
Config

## Motivation

**Chrome Client** 

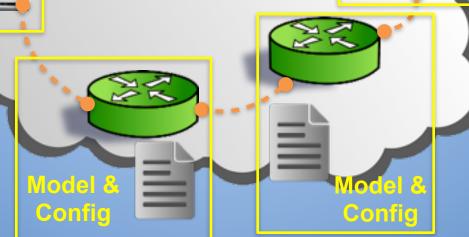
Need detailed network knowledge HW + config

**Apache Server** 



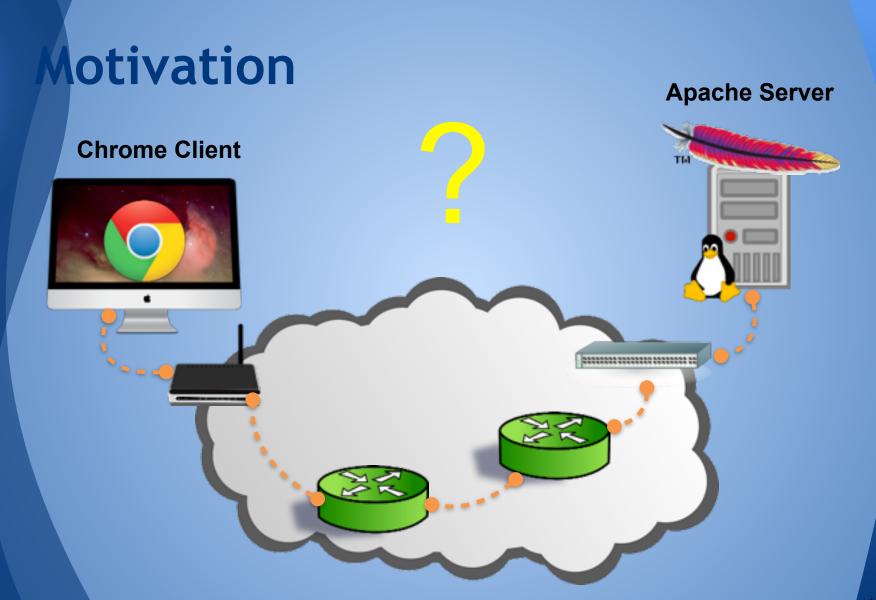
Network Config Analysis

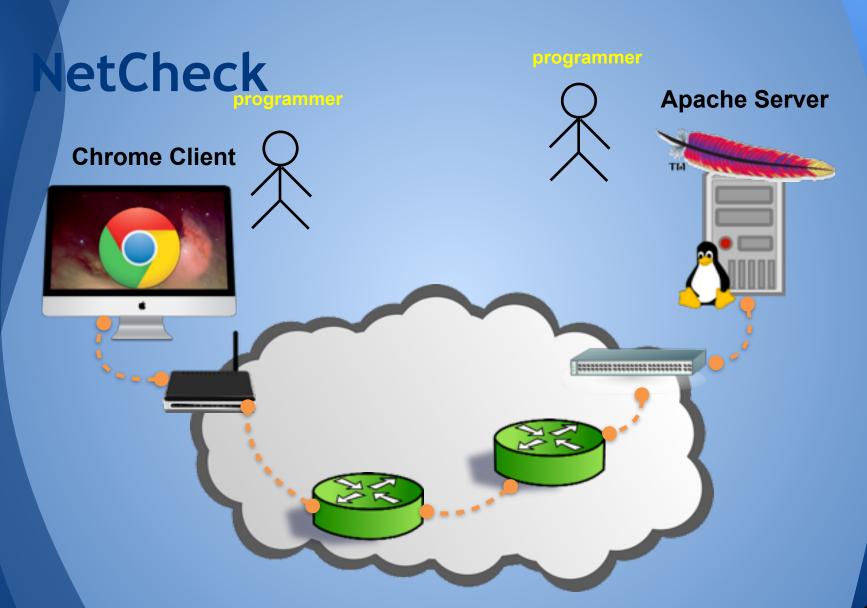
Model & Config



Model &

Config





# **letCheck**

programmer

#### **Chrome Client**





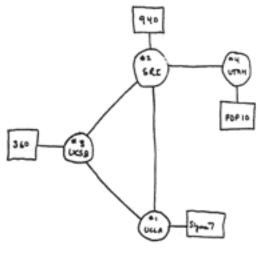
#### programmer



#### **Apache Server**



\*\*\*\*\*\*\*\*



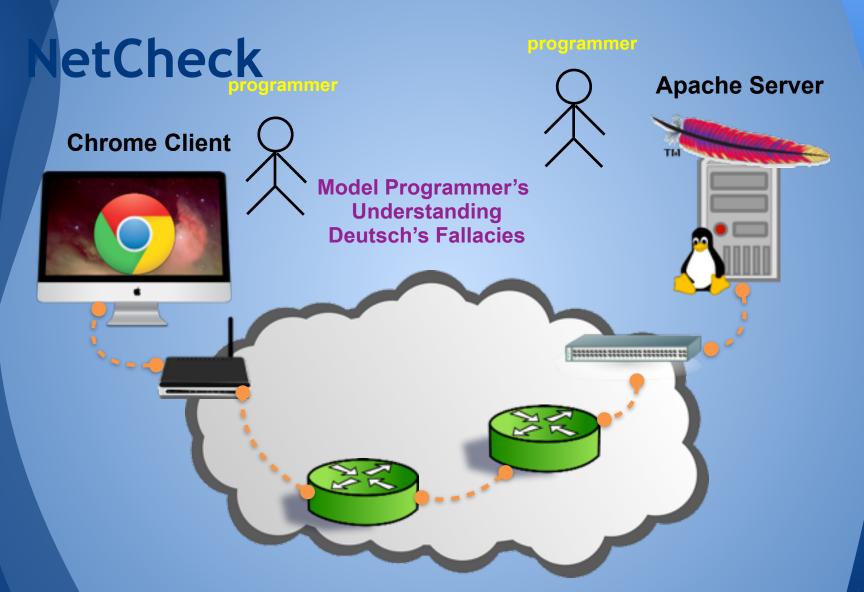
THE ARPA NETWORK

DEC 1969

4 Nobes

Conceptual Sketch of Orignal Internet





# Outline

- Motivation
- NetCheck Overview
- Trace Ordering
- Network Model
- Fault Classification
- Results / Conclusion

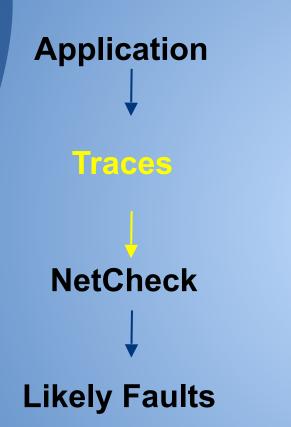
## **VetCheck overview**

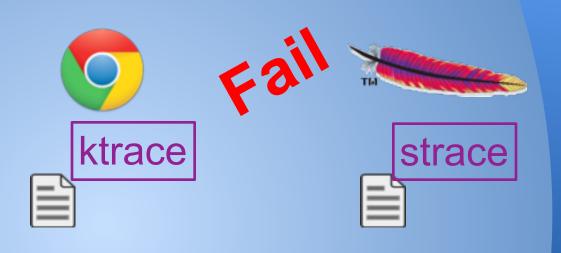
**Application Traces NetCheck Likely Faults** 



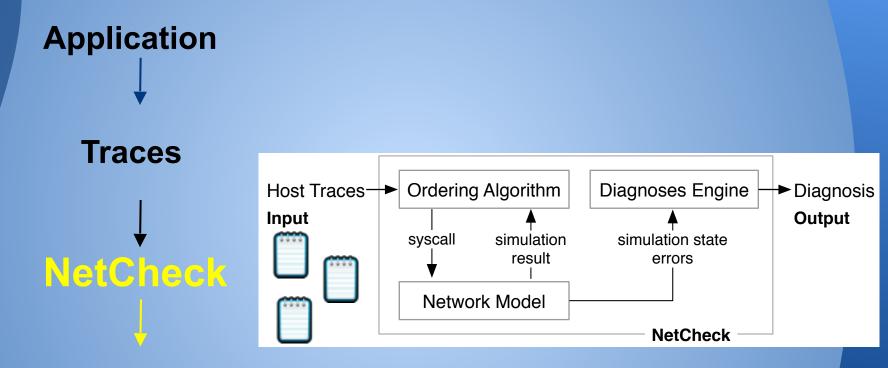


## NetCheck overview





## letCheck overview



**Likely Faults** 

## NetCheck overview

Application

Traces

**NetCheck** 

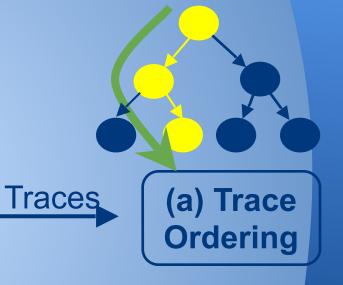
**Likely Faults** 

**Network Configuration Issues** 

twork Configuration Issues uce B fastes to connect to :: I time(s) with an unknown arror This address matches server socket BTP, which was bound to 8.8.8.8:53972 \* [Warning] Poor is an unspecified address '::', which is not portable
\* [Warning] Unable to verify port numbers, which should only happen for the client \* [Warning] Addresses ':: ' and '0.0.0.0' are not the same IP version This address matches server socket B7, which was bound to 0.0.0.0:44126 \* [Warning] Peer is an unspecified address '::', which is not portable [Worming] Unable to verify part numbers, which should only happen for the client [Warning] Addresses '::' and '8.8.8.8' are not the same IP version ailed to connect to 127.8.8.1:55989 1 time(s) because the connection was refused veral nonblocking connects may have failed to connect 1 nonblocking connects from trace D to 127.8.1.1:7777 were never observed to connect numblocking connects from trace C to 127.8.1.1:7777 were never observed to connect raffic from trace D to 239.255.255.258:1988 165 bytes sent, 8 bytes received, 265 bytes lost (188.88%) offic from trace 8 to 239.255.255.258:1988 795 bytes sent, 8 bytes received, 785 bytes lost (188.88%) raffic from trace C to 239,255,255,258;1988 165 bytes sent, 8 bytes received, 165 bytes lost (188.88%) **Problem Detected** The movel factors to process the entire trace [Passible Network Misbehavior] One or more connects failed with an unknown error. This may be due to something filtering connection, for example a firewall. [Ressible Natwork Misbohavior] All 2 datagrams sent from trace D to 210.255.255.256.250 were lest Possible Network Misbehavior] All 6 datagrams sent from trace 0 to 259.255.256.258:1900 were lost Possible Network Misbehavior) All 2 datagrams sent from trace C to 239.255.255.258:1900 were lost If call(s) to getsuckopt, setsockapt, forth, or justh used options which are not currently handled

## Outline

- Motivation
- NetCheck Overview
- Trace Ordering
- Network Model
- Fault Classification
- Results / Conclusion



## Traces

Series of locally ordered system calls

Don't want to modify apps or use a global clock

Gathered by strace, ktrace, systrace, truss, etc.

Call arguments and "return values"

```
socket() = 3
bind(3, ...) Call arguments
lister(3, 1) = 0
accept(3, ...) = 4
recv(4, (HTTP), ...) = 4
Close(4) Return buffer = 0
```

## What we see is this:



### 1. socket() = 3

2. 
$$bind(3, ...) = 0$$

3. 
$$listen(3, 1) = 0$$

4. 
$$accept(3, ...) = 4$$

6. 
$$close(4) = 0$$



1. 
$$socket() = 3$$

2. 
$$connect(3,...) = 0$$

3. 
$$send(3, "Hello", .) = 5$$

4. 
$$close(3) = 0$$

- one trace per host
- local order but no global order

Q: how do we reconstruct what really happened?

## What we want is this

```
A1. socket()
B1. socket()
A2. bind(3, \dots)
                        A3. listen(3, 1)
B2. connect(3,...)
A4. accept(3, ...)
B3. send(3, "Hello", ...) = 5
A5. recv(4, "Hello", ...) = 5
B4. close(3)
A6. close(4)
```







## What we want is this

```
A1. socket()
B1. socket()
A2. bind(3, \dots)
                        A3. listen(3, 1)
B2. connect(3,...)
A4. accept(3, ...)
B3. send(3, "Hello", ...) = 5
A5. recv(4, "Hello", ...) = 5
B4. close(3)
                         = 0
A6. close(4)
```





Goal: find an equivalent interleaving

# Observation 1: Order Equivalence

Node A

$$1. \operatorname{socket}() = 3$$

2. 
$$bind(3, ...) = 0$$

3. 
$$listen(3, 1) = 0$$

4. 
$$accept(3, ...) = 4$$

6. 
$$close(4) = 0$$

- one trace per host
- local order but no global order

Q: how do we reconstruct what really happened?
The socket() calls are not visible to the other side
Some orders are equivalent!



1. 
$$socket() = 3$$

2. 
$$connect(3,...) = 0$$

$$4. \ close(3) = 0$$

# Observation 2: Return Values Guide Ordering

### Node A



2. 
$$bind(3, ...) = 0$$

3. 
$$listen(3, 1) = 0$$

4. 
$$accept(3, ...) = 4$$

6. 
$$close(4) = 0$$

- one trace per host
- local order but no global order

Q: how do we reconstruct what really happened?



1. 
$$socket() = 3$$

2. 
$$connect(3,...) = 0$$

3. 
$$send(3, "Hello", .) = 5$$

4. 
$$close(3) = 0$$

# Return values guide ordering

One valid ordering: all syscalls returned successfully.

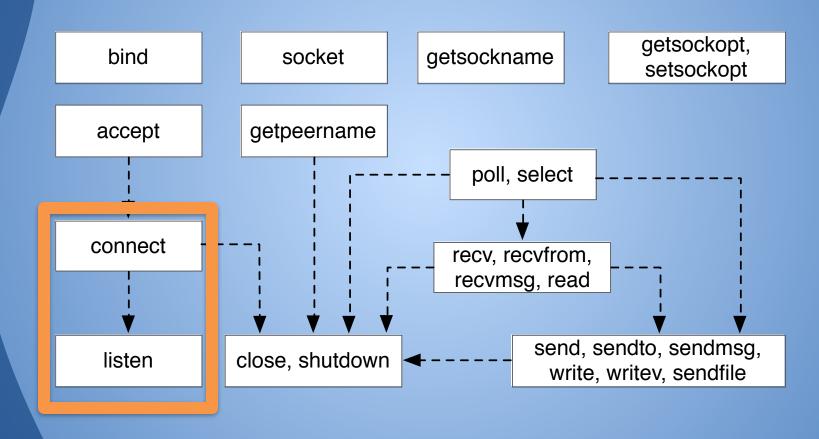
```
A2. bind(3, ...) = 0
B2. connect(3, ...) = -1, ECONNREFUSED
A3. listen(3, 1) = 0
```

A second valid ordering connect failed with ECONNREFUSED.

A call's return value may-depend-on a remote call's action

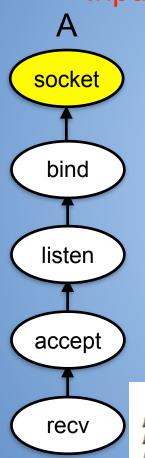
Result indicates order of calls

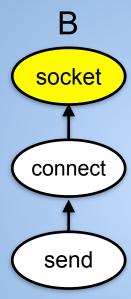
# Deciding call order



full set of may-depend-on relations

#### Input traces





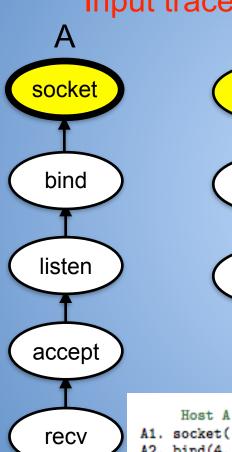
### Algorithm process

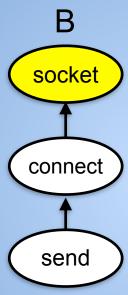
**Output Ordering** 

```
Host A trace:

A1. socket(...) = 4 B1. socket(...) = 3
A2. bind(4, ...) = 0 B2. connect(3, ...) = 0
A3. listen(4, 1) = 0 B3. send(3, "Hello", ...) = 5
A4. accept(4, ...) = 6
A5. recv(6, "Hola!", ...) = 5
```

#### Input traces

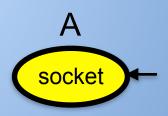




#### Algorithm process

Try socket on host A: accepted

#### **Output Ordering**



```
Host A trace:

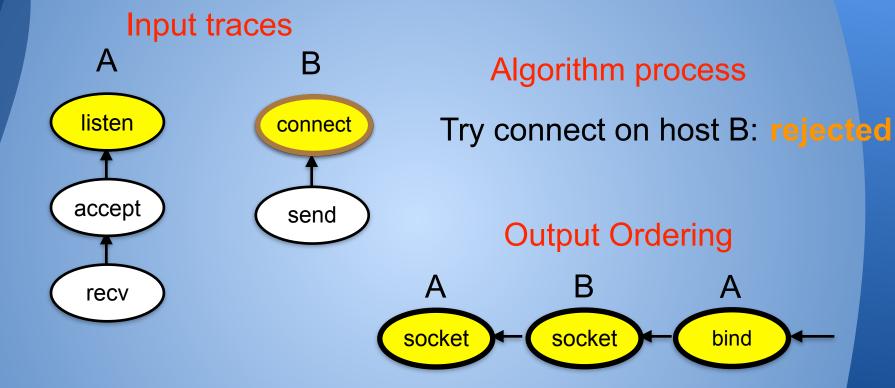
A1. socket(...) = 4 B1. socket(...) = 3

A2. bind(4, ...) = 0 B2. connect(3, ...) = 0

A3. listen(4, 1) = 0 B3. send(3, "Hello", ...) = 5

A4. accept(4, ...) = 6

A5. recv(6, "Hola!", ...) = 5
```



```
Host A trace:

A1. socket(...) = 4 B1. socket(...) = 3

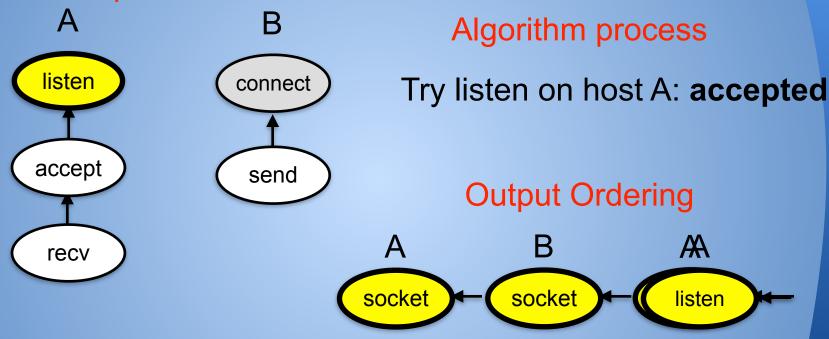
A2. bind(4, ...) = 0 B2. connect(3, ...) = 0

A3. listen(4, 1) = 0 B3. send(3, "Hello", ...) = 5

A4. accept(4, ...) = 6

A5. recv(6, "Hola!", ...) = 5
```

#### Input traces



```
Host A trace:

A1. socket(...) = 4 B1. socket(...) = 3

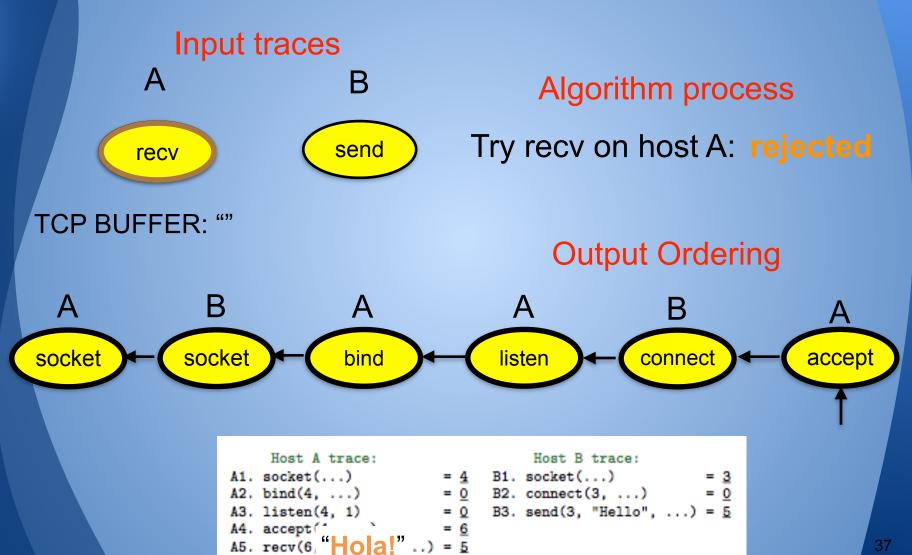
A2. bind(4, ...) = 0 B2. connect(3, ...) = 0

A3. listen(4, 1) = 0 B3. send(3, "Hello", ...) = 5

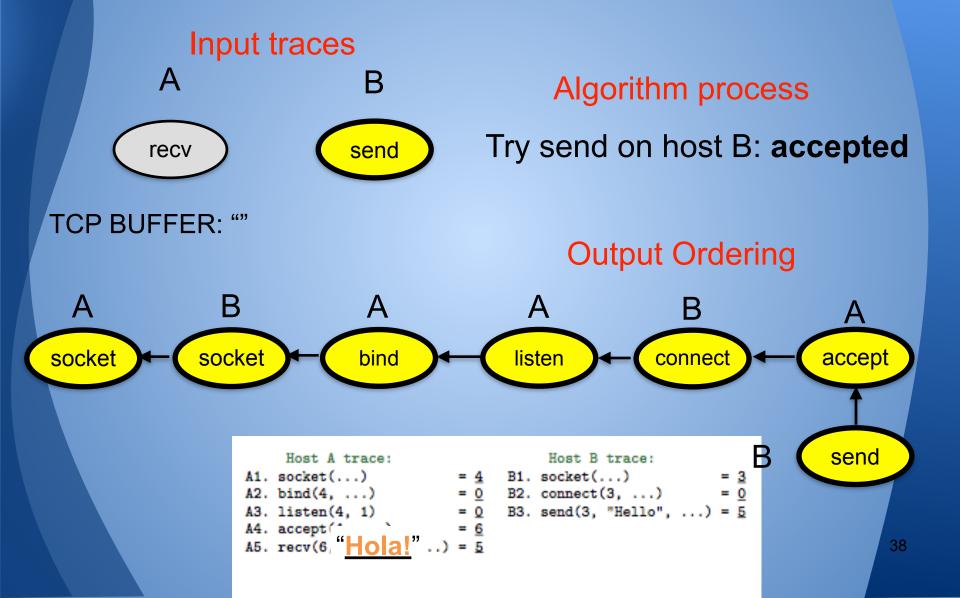
A4. accept(4, ...) = 6

A5. recv(6, "Hola!", ...) = 5
```

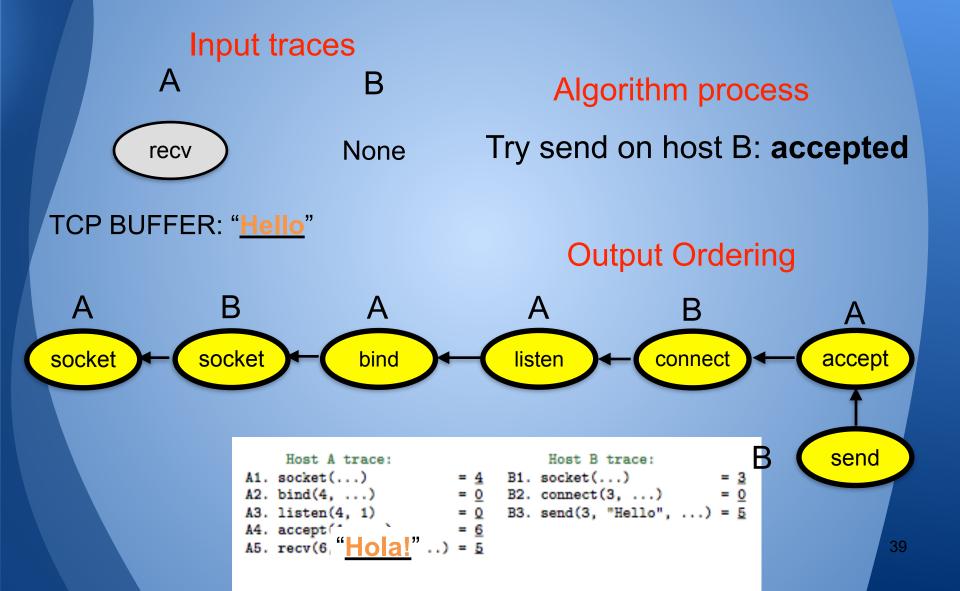
#### rdering Algorithm



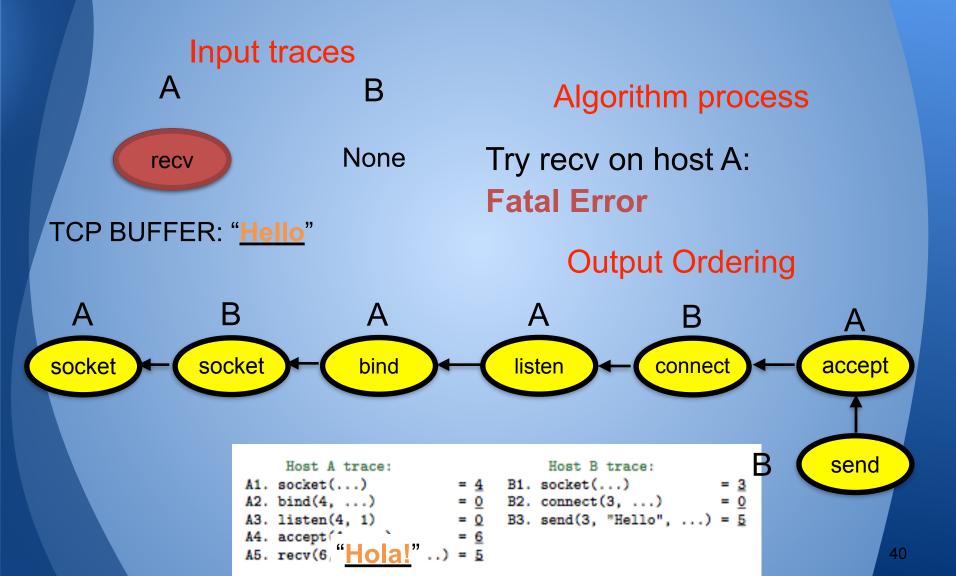
#### **Ordering Algorithm**



### **Ordering Algorithm**

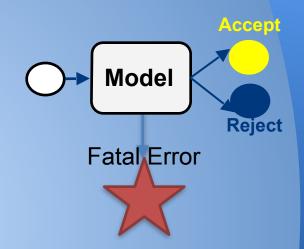


### **Ordering Algorithm**



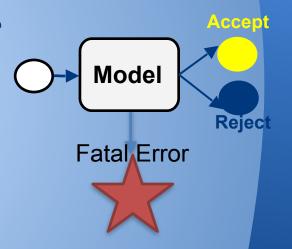
#### Outline

- Motivation
- NetCheck Overview
- Trace Ordering
- Network Model
- Fault Classification
- Results / Conclusion



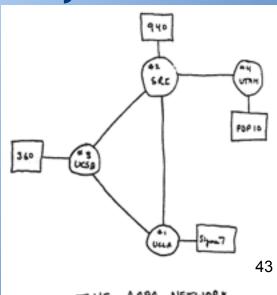
#### **Network Model**

- Simulates invocation of a syscall
  - datagrams sent/lost
    - reordering / duplication is notable
  - track pending connections
  - buffer lengths and contents
    - send -> put data into buffer
    - recv -> pop data from buffer
- Simulation outcome
  - Accept → can process (correct buffer)
  - Reject → wrong order (incomplete buffer)
  - Permanent reject → abnormal behavior (incorrect buffer)



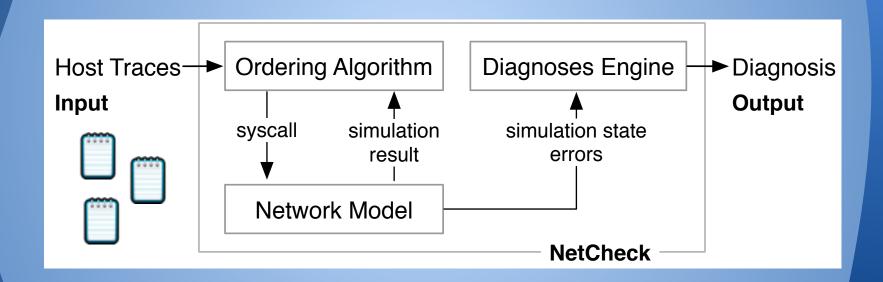
#### **Network Model**

- Simulates invocation of a syscall
- Capture programmer assumptions
  - Assumes a simplified network view
    - Assume transitive connectivity
    - Little, random loss
    - No middle boxes
  - Assume uniform platform
    - Flag OS differences



# ow Model Return Values Impact race Ordering

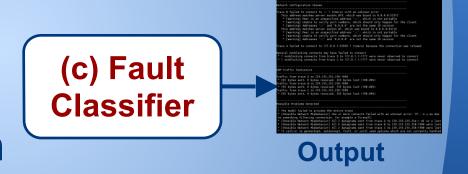
Blackbox Tracing mechanism



Trace Ordering: linear running time (total trace length) \* number of traces

#### Outline

- Motivation
- NetCheck Overview
- Trace Ordering
- Network Model
- Fault Classification
- Results / Conclusion



#### Fault Classifier

- Goal: Decide what to output
- Problem: Show relevant information
- Fault classifier: global (rather than local) view
  - uncovers high-level patterns by extracting low-level features
    - Examples: middleboxes, non-transitive connectivity, MTU, mobility, network disconnection
    - All look like loss, but have different patterns in the context of other flows

#### Fault Classifier

- Options to show different levels of detail
- Network admins / developers
  - detailed info
- End users
  - Classification
  - Recommendations

**Network Configuration Issues** 

```
ace B failed to connect to :: I time(a) with an unknown error
 This address matches server socket BT9, which was bound to 0.0.0.2:53973
   '[Warming] Peer is an unspecified address '::', which is not pertable
    [Warning] Unable to verify port numbers, which should only happen for the client
  * [Warming] Addresses '!!" and '0.0.0.0" are not the same IP version
 This address matches server socket 67, which was bound to 8.8.8.8:44126
   [Warning] Peer is an unspecified address ":: ", which is not pertable
    [Warming] Unable to verify part numbers, which should only happen for the client
                             and '8.8.8.8' are not the same IP version
                 onnect to 137.8.8.1:55989 1 time(s) because the connection was refused
      nonblocking connects may have failed to connect
   numblocking connects from trace D to 127.8.1.1:7777 were never
raffic from trace D to 259.255.255.258:198#
165 bytes sent, 0 bytes received, 265 bytes lost (180.00%)
raffic from trace 8 to 239.255.255.258:1988
 795 bytes sent, 8 bytes received, 785 bytes lost (188.88%)
     from trace C to 239,255,255,258:1988
```

#### Problem Detected

- \* (Possible Nation: Misbehavior) One or more connects failed with an unknown error. This may be do to samething filtering connection. For example a firewall.
- \* [Pussible Natura Misbehavier] All 3 datagrams sent from trace D to 350.355.355.356:1989 were lest
  \* [Pussible Natura Misbehavier] All 6 datagrams sent from trace D to 350.355.355.356:1987 were lest
- \* [Possible Network Misbehavior] All E datagrams sent from trace E to 139.255.255.256.1569 were list \* [Possible Network Misbehavior] All 2 datagrams sent from trace C to 139.255.255.25611569 were list \* If call(s) to getsickopt, setsockopt, font], or lict( used options which are not currently handled.

#### Outline

- Motivation
- NetCheck Overview
- Trace Ordering
- Network Model
- Fault Classification
- Results / Conclusion

# Evaluation: Production Application Bugs

- Reproduce reported bugs from bug trackers (Python, Apache, Ruby, Firefox, etc.)
  - A total of 71 bugs
  - Grouped into 23 categories
    - Virtualization incurred/portability bugs
    - SO\_REUSEADDR behaves differently across OSes
    - accept inherit O\_NONBLOCK
    - •••
  - Correct analysis of >95% bugs

## Faults Observed Network

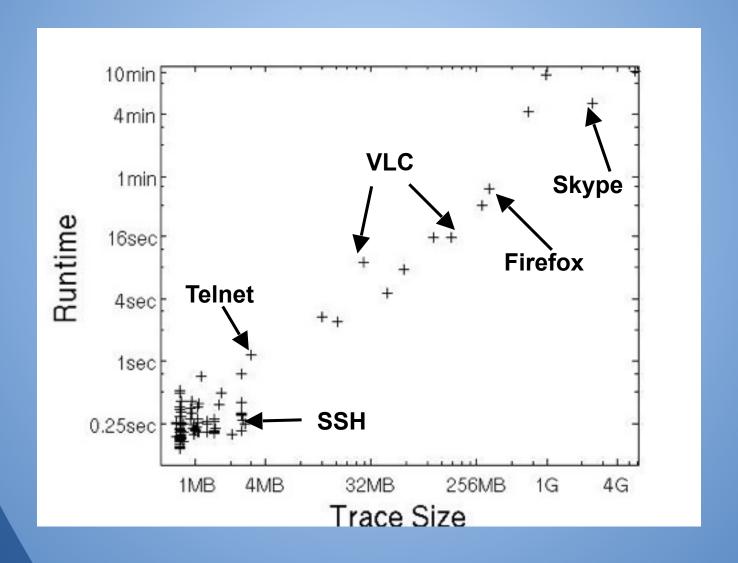
- Twenty faults observed in practice on a live network
  - MTU bug
    - Intermediary device
  - Port forward
    - Traffic sent to non-relevant addresses
  - Provide supplemental info
    - packet loss
    - buffers being closed with data in
  - 90% of cases correctly detected

## General Findings in Practice

- Middle boxes
  - Multiple unaccepted connections
    - client behind NAT in FTP
- TCP/UDP
  - non-transitive connectivity in VLC
- Complex failures
  - VirtualBox send data larger than buffer size
  - Pidgin returned IP different from bind
  - Skype NAT + close socket from a different thread
- Used on Seattle Testbed <u>seattle.poly.edu</u>



#### NetCheck Performance Overhead



#### Conclusion

Built and evaluated NetCheck, a tool to diagnose network failures in complex apps

- Key insights:
  - model the programmer's misconceptions
  - relation between calls → reconstruct order
- NetCheck is effective
  - Everyday applications & networks
  - Real network / application bugs
  - No per-network knowledge
  - No per-application knowledge

Try it here: https://netcheck.poly.edu/

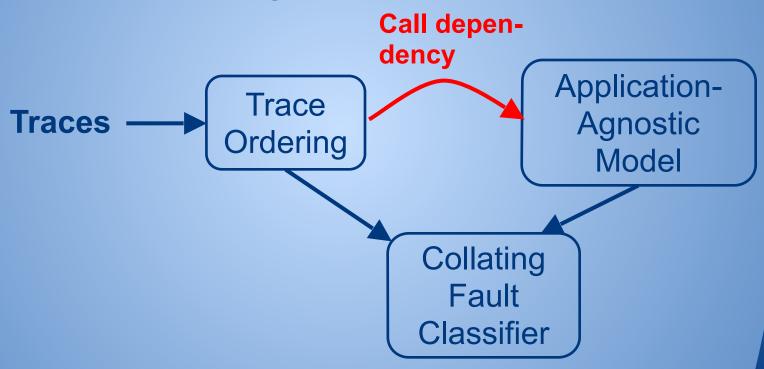
## Backup slides.

#### What is NetCheck?

- No app- or network-specific knowledge
- No modification to apps/infrastructure
- No synchronized global clock
- Blackbox Tracing mechanism (eg, strace)
  - Reconstruct a plausible total ordering of syscall traces from multiple hosts
  - Uses simulation and captured state to identify network related issues
  - Map low-level issues to higher-level characterizations of failure

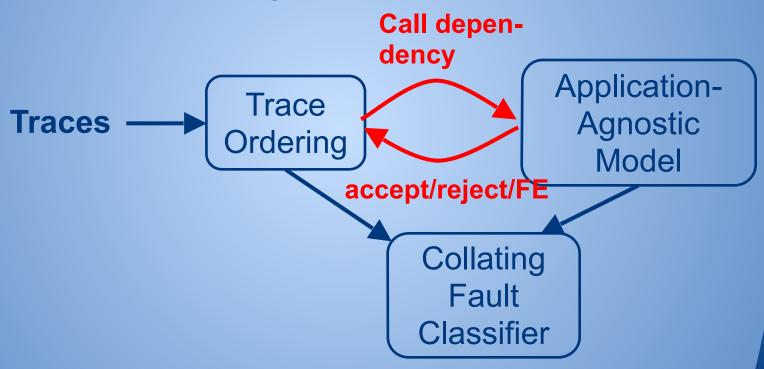
## **Diagnosis Model**

Blackbox Tracing mechanism



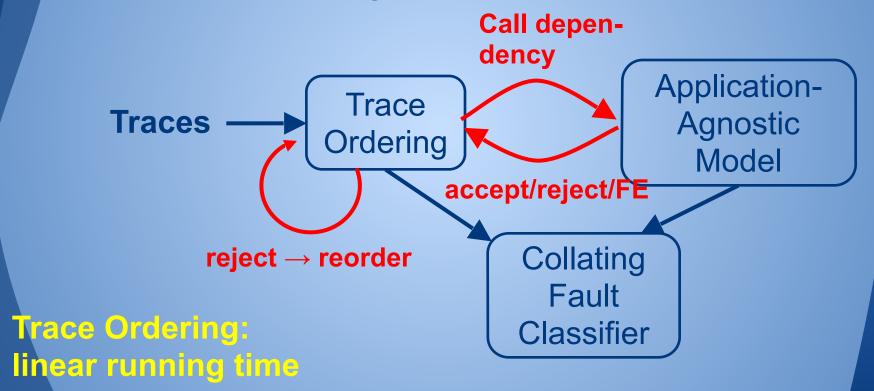
## Diagnosis Model

Blackbox Tracing mechanism



## Diagnosis Model

Blackbox Tracing mechanism



## Pseudocode and Analysis

```
1. push trace t<sub>0</sub> in stack s<sub>0</sub>, ..., trace t<sub>n-1</sub> in stack s<sub>n-1</sub>
2. while (s<sub>0</sub>, ..., s<sub>n-1</sub>) not empty:
3. q = peek_stack(s_0, ..., s_{n-1}); q.sort(priority)
                                             Best case: O(1)
      if q empty: raise FatalError
5.
                                          Worst case: O(n)
      i<sub>i</sub> = q.dequeue();
6.
7.
      outcome = model_simulate(i;)
                                                                  Overall:
8.
      if outcome == ACCEPT:
                                                         Best case O(L)
        ordered_trace.push(si.pop()); break Worst Case O(n*L)
9.
10.
       elif outcome == REJECT:
                                        pass
11.
       elif outcome == FatalError: raise FatalError
```

## Pseudocode and Analysis

```
1 push trace t<sub>0</sub> in list s<sub>0</sub>, ..., trace t<sub>n-1</sub> in list s<sub>n-1</sub>
2. while (s_0, \ldots, s_{n-1}) not empty:
3. q = peek_stack(s_0, ..., s_{n-1}); q.sort(priority)
   while True:
5. if q empty: raise FatalError
      i_i = q.dequeue();
6.
7.
      outcome = model_simulate(i;)
     if outcome == ACCEPT. Accept → Traverse
8.
        ordered_trace.push(s<sub>i</sub>.pop()); break
9.
     elif outcome == REJECT: continue Reject → Backtrack
10.
      elif outcome -- FatalError: raise FatalError
```

## NetCheck input

```
Node A

1. socket() = 3
2. bind(3, ...) = 0
3. listen(3, 1) = 0
4. accept(3, ...) = 4
5. recv(4, "Hello", ...) = 5
6. close(4) = 0
```

```
Node B
1. socket() = 3
2. connect(3,...) = 0
3. send(3,"Hello",...) = 5
4. close(3) = 0
```

## NetCheck input

```
Node A

1. socket() = 3
2. bind(3, ...) = 0
3. listen(3, 1) = 0
4. accept(3, ...) = 4
5. recv(4, "Hello", ...) = 5
• 6. close(4) = 0
```

#### Node B

```
1. socket() = 3
2. connect(3,...) = 0
3. send(3, "Hello",.) = 5
4. close(3) = 0
```

### connect depends on listen

```
Order 1 A1 bind(3, ...) = 0
         A2 listen(3, 5) = 0
         B1 connect(3, ...) = 0
Order 2 A1 bind(3, ...) = 0
         B1 connect(3, ...) = -1 ECONNREFUSED
         A2 listen(3, 5) = 0
Order 3 B1 connect(3, ...) = -1 ECONNREFUSED
         A1 bind(3, ...) = 0
         A2 listen(3, 5) = 0
```

## **Example Rules**

- Middle boxes
  - Multiple unaccepted connections
    - ⇒ client behind NAT in FTP
  - Missing connect on accepted connections → server behind NAT or port forwarding
  - Multiple connect non-standard failure → firewall filtering connections
  - Multiple connect to listening address get refused
  - Multiple non-blocking connect failure
  - Traffic sent to non-relevant addresses → NAT or 3rd party proxy/traffic forwarding

## Example fault classifier rules

- Middle boxes
  - Multiple unaccepted connections
    - ⇒ client behind NAT in FTP
  - Missing connect on accepted connections → server behind NAT or port forwarding
  - Traffic sent to non-relevant addresses → NAT or 3rd party proxy/traffic forwarding
- TCP
  - select/poll timeout
  - send data after connection closed

## Example rules (cont.)

- UDP
  - datagram sent/lost per connection
  - high datagram loss rate
    - ⇒ non-transitive connectivity in **V**(C)

- Misc
  - apps send data larger than default OS buffer size
     ⇒ bug report from VirtualBox bug tracker
  - returned IP different from bind
     ⇒ simultaneous net disconnect/reconnect in Pidgin
  - Skype attempted to close socket from a different thread

## Evaluation: Everyday Applications

- FTP
  - All reverse connections from server lost
    - Client behind NAT
- Pidgin
  - getsockname returns different IP
    - Client poor connection results in IP changes
- Skype
  - Poor call quality, msg drop
    - Network delay, NAT
    - Skype closes socket from different thread
- VLC
  - Packet loss
    - Non-transitive connectivity issue