



A Semantic Framework for Data Analysis in Networked Systems

Arun Viswanathan, Alefiya Hussain, Jelena Mirkovic, Stephen Schwab, John Wroclawski

USC/Information Sciences Institute

Data Analysis in Networked **Systems**



type-inviol_ mag-music(1200-00177-010-10170); arch-c0000008 syscall-11 success-yes exit-0 a0-76c2b0c75000 a1-3000 a2-0 a0-322030553838244f imag-0 gpdm=41445 pid=31855 auid=4291957205 uid=1000 gld=1004 e01d=1000 suid=1060 fsuid=1000 e01 u=3003 agit=1000 fmgid=13000 fmg-(0100) fmg=420497205 comm="df" exc="/min/df" key=(ut1)

Land3 spil-1000 fegid-1000 ffgid-1000 ffg-(none) sev-4254957285 come."df" eke-"/bin/df" key-(null) typenyscalt mod-audit(1295538727.m38:75171): arch-cug0000e syscalt-2 Struess-yee exit-3 a0-La2c636 al-6 a2-La3c670 a3 1000 frazid-1000 fragid-1005 prid-31365 null-429097295 crange"/f" aver//bin/df" key=(null) type=%178 hsg-audit(129538732.h0.75171): two-fn temer//bin/df" key=(null) type=%178 hsg-audit(129538732.h0.75171): two-fn temer/lab/df" key=(null) type=%178 hsg-audit(129538732.h0.75171): two-fn temer/lab/df" key=(null) type=%178 hsg-audit(129538732.h0.75171): two-fn temer/lab/localc/en US.at f8/LC T0ENTIFICATION" inode-151655 dev type=5%27611 mg-audit(129538732.h0.75172): arch-ch0 sx1900 as-Laarnag incerce puto-31855 pid-31965 pid-31965 co 1 al-2 stressed pare-33355 pid-31955 nid=31965 nid=420867795 co 0 fraid-2006 fid-1000 fid-1000 sg-174f44

13-2 storged pp:0-31355 pid-31926 Auid-4204067295 bi

Did my

experiment

run as

expected?

16

1/4

.168

Mozil

192.16

la

107

2.160.7

(compat

2.168.1.51

PATH mag-could+t12-PUSIE (1)

and and and the

6100644

M LLy-(nnex) ves-4244657295 rdmm-df (ntwo) sep=40+4887295 rdmm="df" exec*/bin/df" key=(mrtt) malt(128538752,638;75:74): arcm=c0600085c stacatt=3 success=yes exit=3 a0=3 a1=175 d2=1 a3=2 items=0 95% audie=229.c622945 ord=1000 gd=1000 suid=1600 fubid=1600 cyid=1000 syid=1000 fagid=1000 fagid=1000 -175 at-8 at-175 at-EVERYSCALL MSG 1955 girl=31958 pinesises automation of " gree"/bin/dr" koy-(hult)

cs=404065/200 comm="df" unc="/bin/df" kgy=ibult) eng-dmdie(100451072.536:75175); arch=4000000e syscalt=2 success-yes exit=3 a0-7fs2600149eB al=0 a2-0 a3-i teased prid=1005 pd=10066 arch=439400705 uld=1000 01d=1000 exit=4000 exit=1000 fruid=1000 exit=1000 s fagad-1000 ou-audit (1209538732,636

Is there

any evidence

of a known

attack?

2.168.1.51 [11/scp/2009:08:33:18 La/4.0 (compatible; MSIE 5.5; Win32)"

: Win32)"

5; Win32)"

listy: ≪/lin/geon//geon/-motules.coche" inodex132375 dev+07:08 nod

vacali=5 Success=yCs Cx1t=0 d0=3 al=7hft437317c8 B2=/t1t4 4567245 uid=1690 gid=1060 cuid=1090 suid=1000 [suid=1000



W1N32)" [30:36 -0700] "GET /HNAP1/ HTTP/1.1" 404 288 "-" "Moz Min32)" "30:40 -0700] "GET /TEADevInfo/ HTTP/1.1" 404 293 "."

.5; Win32). 30:43 -0700] "GET / HTTP/1.0" 200 980 "." "Mozitla/4 30:43 -0700] "GET /HNAP1/ HTTP/2 1" :08:33:12 -07001

09:00:33:15 -0700 MSIE 5.5; Win32)"

te: MSIE 5.5; Win32)"
p/2009:08:33:18 ~0700] "GET / HTTP/1.0" 200 980 "." "Mozilla/4 2.100.1.31 La/4.0 (compatible; MSIE 5.5; Win32)" La/4.0 (compatible; MSIE 5.5; Win32)" La/4.0 (compatible; MSIE 5.5; Win32)" "GET /HMAP1/ HTTP/1.1" 404 288 "." "Moz 2.168.1.51 La/4.0 (compatible; MGIE 5.5; WIN32)" 2.168.1.51 [11/Sep/2009:00:39:59 +0700] "GET /TEADevInfo/ HTTP/1.1" 404 293 "."

03/07-13:32:12.330048 192.168.2.53 -> 192.168.2.51 ICMP TTL:64 T05:0x0 ID:36945 IpLen:20 DgmLen:84 Type:0 Code:0 ID:20247 Seq:194 ECHO REPLY 3

ICNP TTL:64 T05:0x0 ID:0 1010 ID:00100 Type:8 Code:0 ID:20247 Seq:195 ECHO

[**] [1:364:5] ICMP PING [**] [Classification: Misc activity] [Prio [Classification: 334038 192.168.2.5]

[Classification: Misc activity] [P 83/07.13:32:13.334038 192.168.2.51 93/07.15:64 TOS:0x0 ID:0 Iptmn.20 03/07-13:32:13,334038 192,168,2,5 1000 TTL:64 TOS:0x0 TD:0 IPLEN:20 ICMP .s Code:0 ID:20247 Sen:10

10. 11"

201

10

110

10

20

10

10

Packet Dumps

TP'

TP.

[**] [1:366:7] ICMP PING *NIX [**]
[Classification: Misc activity] [Priority: 3]
[Classification: 334838 192.168.2.51 -> 192 ***

03/07-13:32:13.3340:56 197.168.2.51 -> 192.168. 03/07-13:32:13.3340:56 197.168.2.51 -> 192.168. 10:0 TTL:64 TOS:0x0 ID:0 IpLen:20 DgmLen:84 DF 1CMP TTL:64 TOS:0x0 ID:20247 Seq:195 ECHO

[Classification: Misc activity] [Priority: 3] [Classification: Misc activity] [Priority: 3] 03/07-13:32:13.334036 192.160.2.51 -> 192.168.2.53 03/07-11:64 T05:0x0 ID:0 IpLen:20 DomLen:84 DF

Alerts

Why did

failure X

happen?

[**] [1:368

Type; 5

e۴

88: 44:

6.6

e ^g

38-2515BH

- 1

92.108.2.53

DE

Is my

hypothesis

validated?

2226- 1/1/1

\$777A 2228**

32778. 4.2.33770

33783

A7

150

A 159 1e

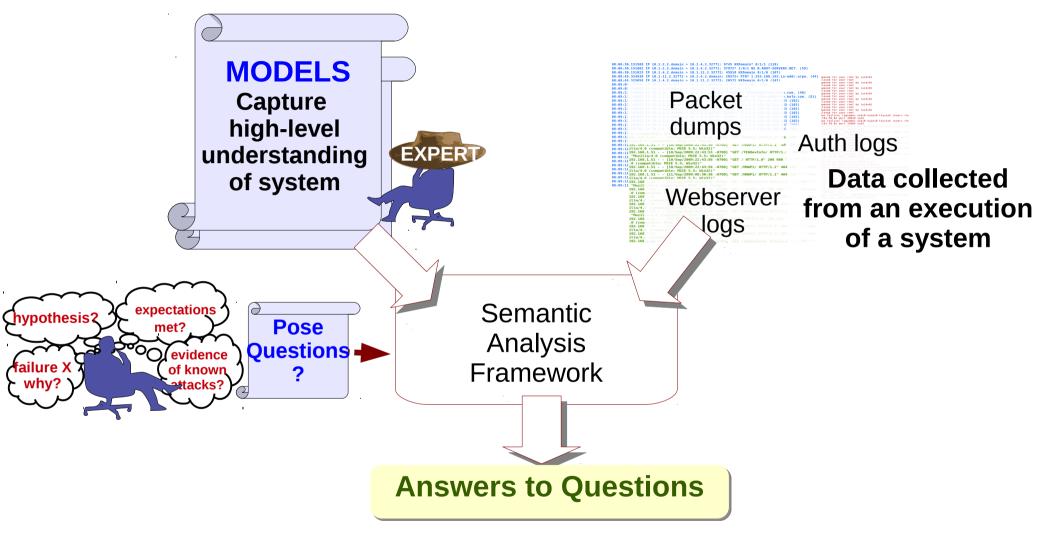
.30

0

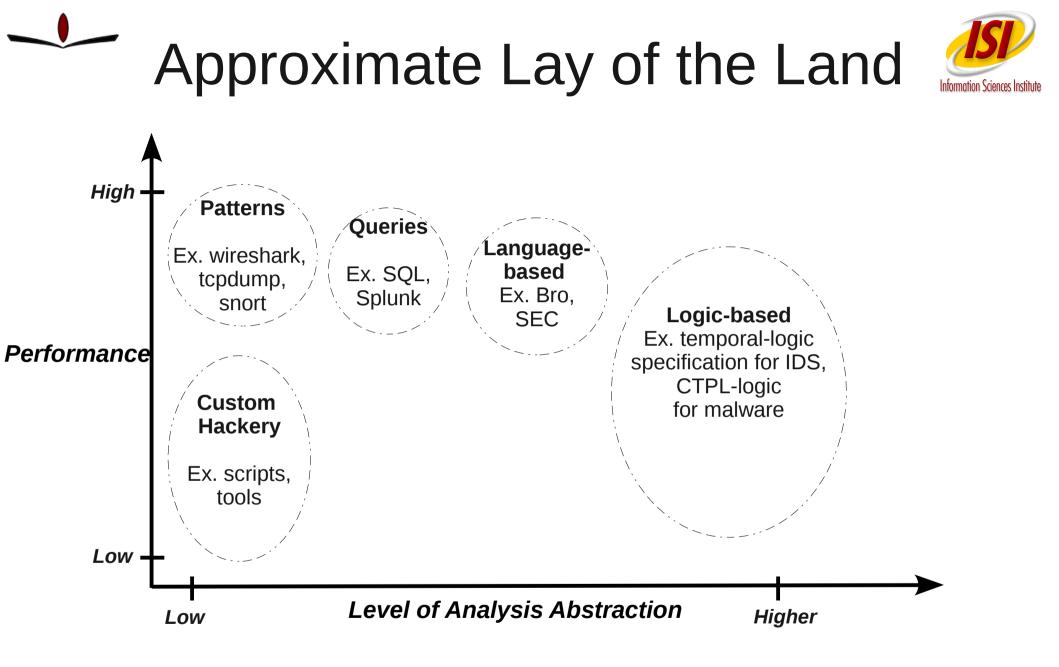
0

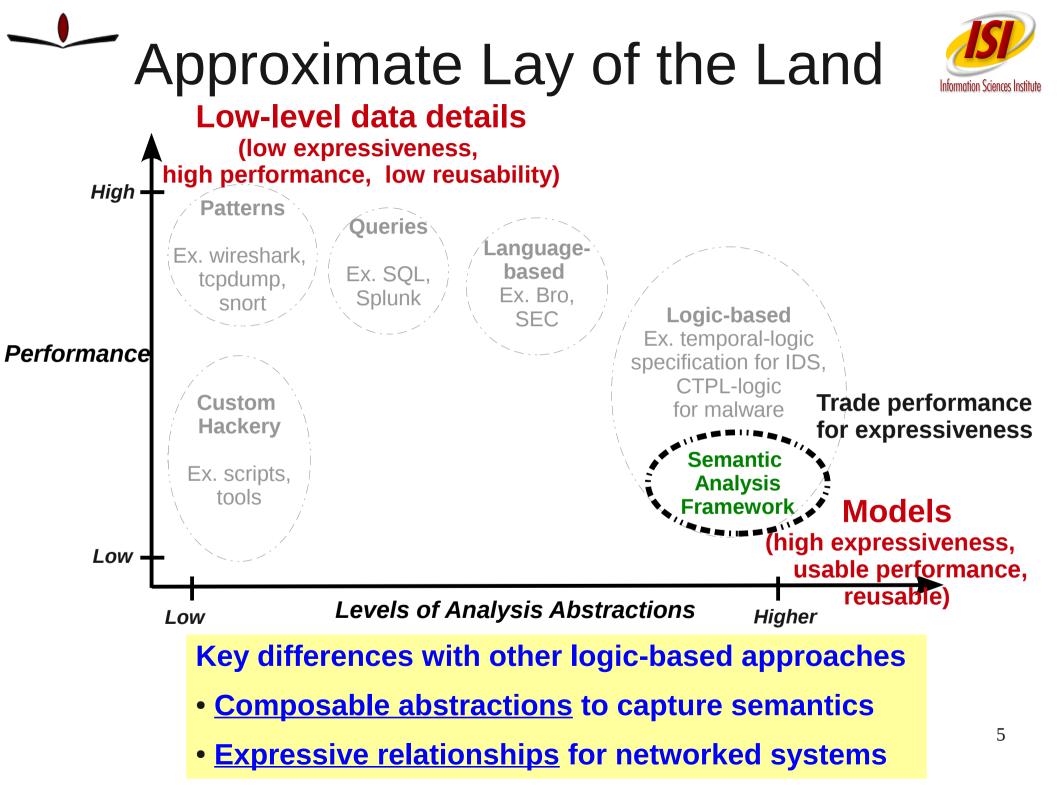
Our Semantic Approach



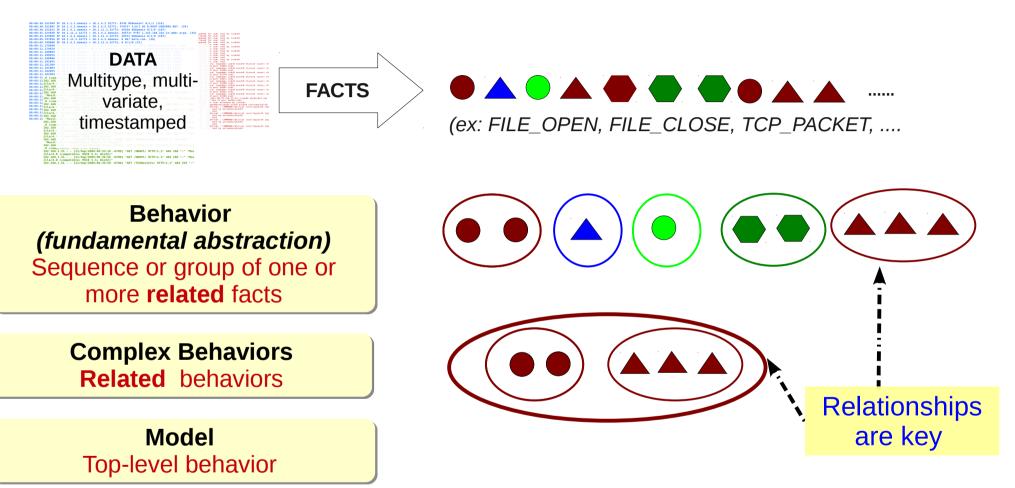


Models drive analysis over data!





Basics of our Modeling Approach



Models encode higher-level system semantics!

Relationships in the Modeling Language



Temporal Relationships

Causality/Ordering Eventuality Invariance Synchrony/Timing

A file open eventually leads to a file close

HTTP and FTP flows are

concurrent.

Temporal Operators

FILE_OPEN ~> FILE_CLOSE

Concurrent Relationships

Parallelism Overlaps

Logical Relationships

Combinations Exclusions

Dependency relationships b/w data attributes Experiment **either** succeeds **or** fails

or fails

File open and file close are behaviors **related by their filename**. Interval Temporal Operators

HTTP_FLOW **olap** FTP_FLOW

Logical Operators

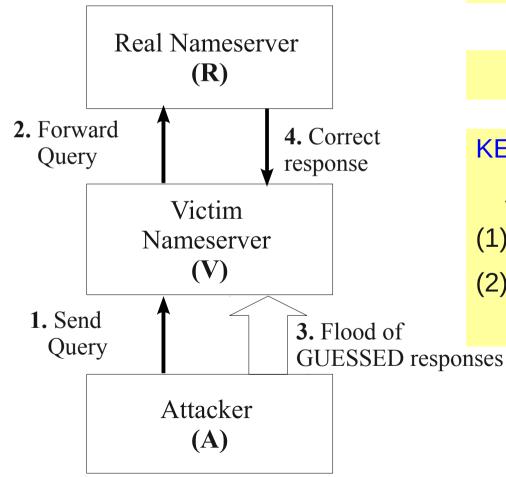
EXPT_SUCCESS **xor** EXPT_FAIL

FILE_CLOSE.name = FILE_OPEN.name



Cache Poisoning Behavior

Cache Poisioning Behavior (DNS Kaminsky)



Objective: Attacker poisons the victim's DNS cache.

Steps 1-4 keep running in a loop.

KEY ISSUES

Attacker fails to poison cache due to

(1) Race conditions with real nameserver.

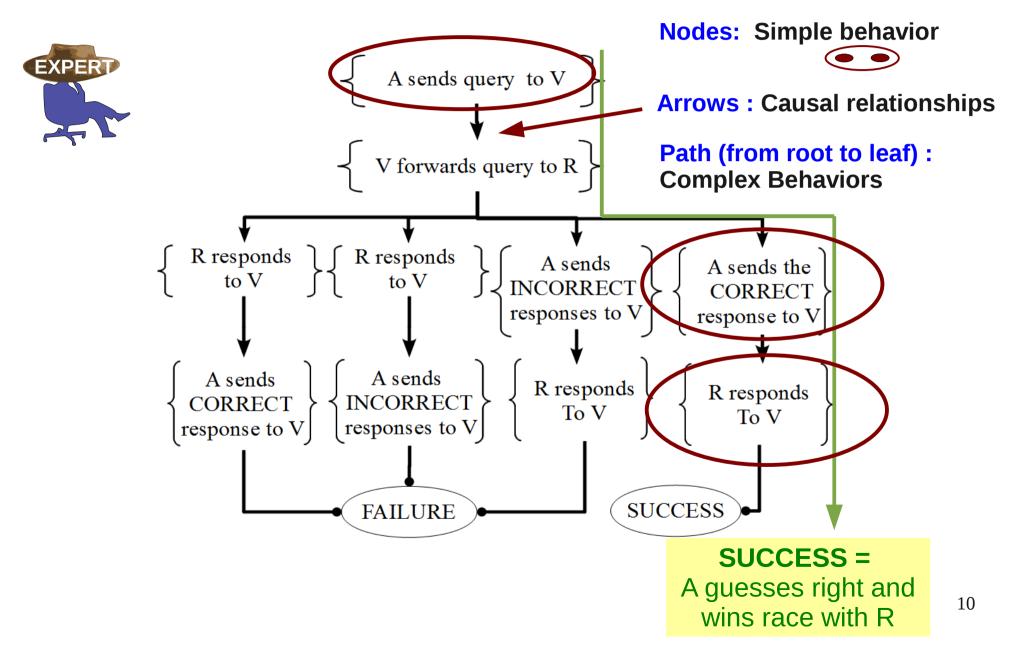
(2) Incorrectly GUESSED responses.

Analysis using typical approach 🌽

```
08:06:30.151588 IP 10.1.2.2.domain > 10.1.4.2.32771: 8745 NXDomain* 0/1/1 (118)
08:06:30.151602 IP 10.1.2.2.domain > 10.1.4.2.32771: 37972* 1/0/1 NS B.ROOT-SERVERS.NET. (59)
08:06:30.151823 IP 10.1.4.2.domain > 10.1.11.2.32772: 45918 NXDomain 0/1/0 (107)
08:08:45.324938 IP 10.1.11.2.32772 > 10.1.4.2.domain: 20572+ PTR? 1.253.168.192.in-addr.arpa. (44)
08:08:45.325058 IP 10.1.4.2.domain > 10.1.11.2.32772: 20572 NXDomain 0/1/0 (107)
08:09:09.787858 IP 10.1.11.2.32772 > 10.1.4.2.domain: 0 NS? bofa.com. (26)
08:
                           Tricky to analyze
08:
08:
                                                                                    (51)
08:
       Requires Expertise.
08:
08:
      Too many random values in the data to extract
88:
08:
08:
      using simple patterns.
08:
08:
08:
       Race conditions (timing issues) are hard to
08:
08:
      debug over 10's of thousands of packets.
08:
                                                                                   om. (52)
08:
08:
       Many ways to fail.
08:
08:
08:
08:09:11.185802 IP 10.1.6.3.domain > 10.1.4.2.32778: 2228*- 1/1/1 A 159.16.126.233 (103)
08:09:11.186301 IP 10.1.6.3.domain > 10.1.4.2.32778: 2228*- 1/1/1 A 159.16.126.233 (103)
08:09:11.186551 IP 10.1.6.3.domain > 10.1.4.2.32778: 2228*- 1/1/1 A 159.16.126.233 (103)
08:09:11.186812 IP 10.1.6.3.domain > 10.1.4.2.32778: 2228*- 1/1/1 A 159.16.126.233 (103)
08:09:11.187051 IP 10.1.6.3.domain > 10.1.4.2.32778: 2228*- 1/1/1 A 159.16.126.233 (103)
```

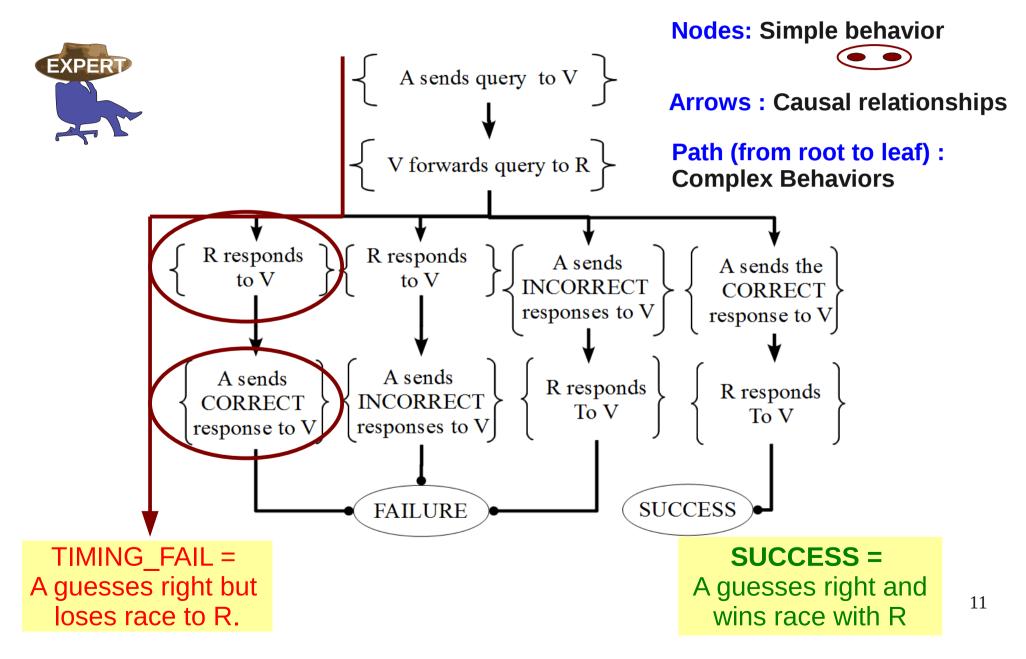


Model of Behavior

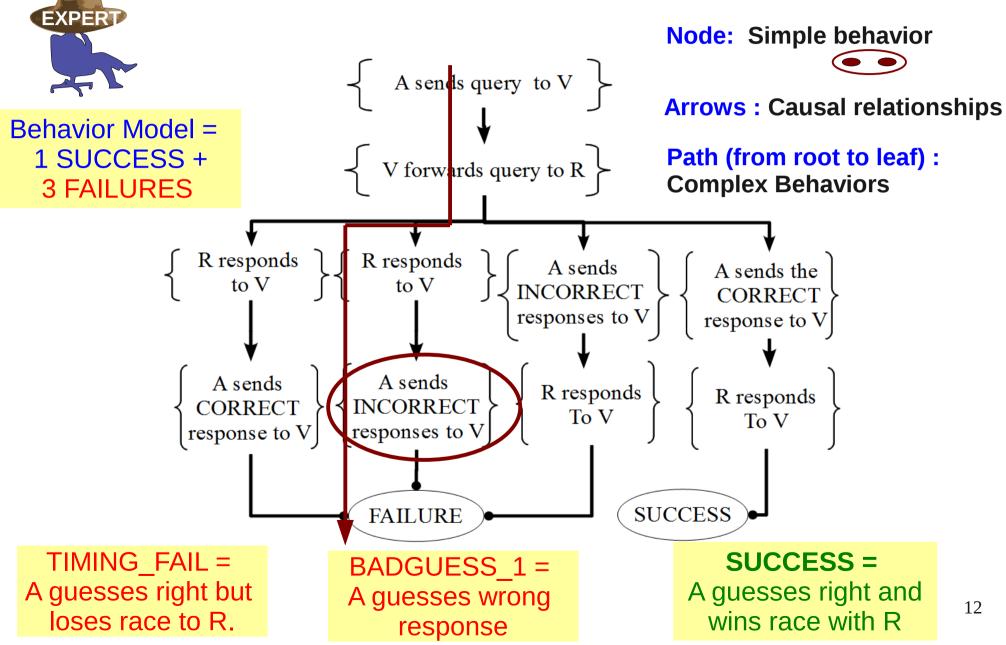




Model of Behavior



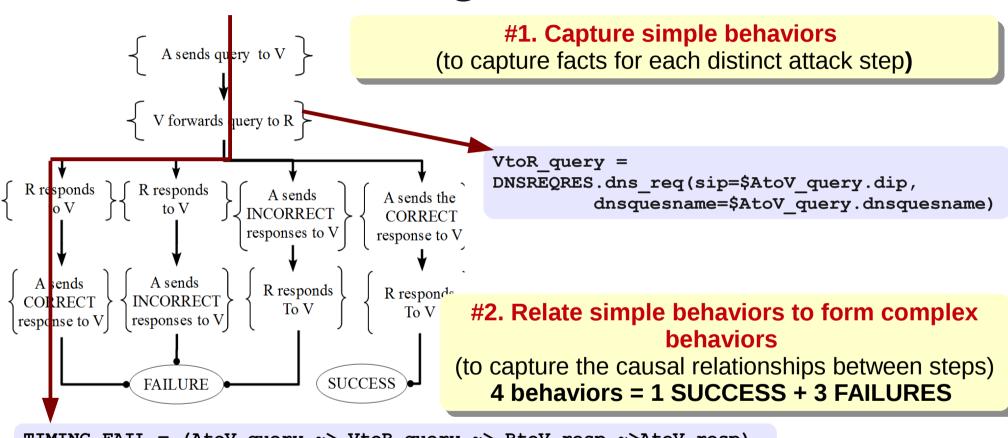
Model of Behavior







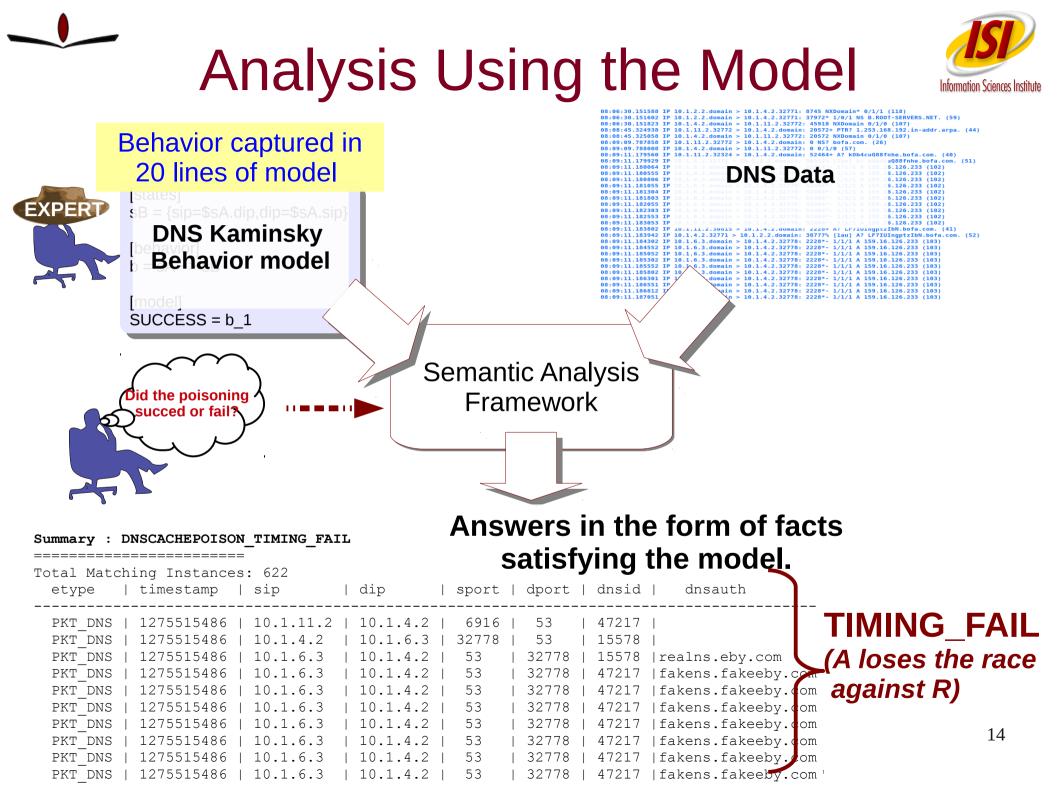
Encoding the Model



TIMING_FAIL = (AtoV_query ~> VtoR_query ~> RtoV_resp ~>AtoV_resp)

#3. Define Behavior Model

(assertion to capture users understanding of system operation)





Current Implementation and Performance

- Prototype algorithm for applying models over data.
- Algorithm performance
 - O(N²) worst-case performance
 - Straight-forward

Analysis Framework

- Written in Python
- SQLite-based storage backend
- Scalability and performance issues are under active investigation.

Applicability



- Broad range of event-based modeling in networked systems
- More examples in paper
 - Modeling hypotheses
 - Ex. Validating DoS detection heuristics over traces
 - Modeling a security threat
 - Ex. Model of a simple worm spread over IDS logs
 - Modeling dynamic change
 - Ex. Model of changes in traffic rate due to attack.

Future Work



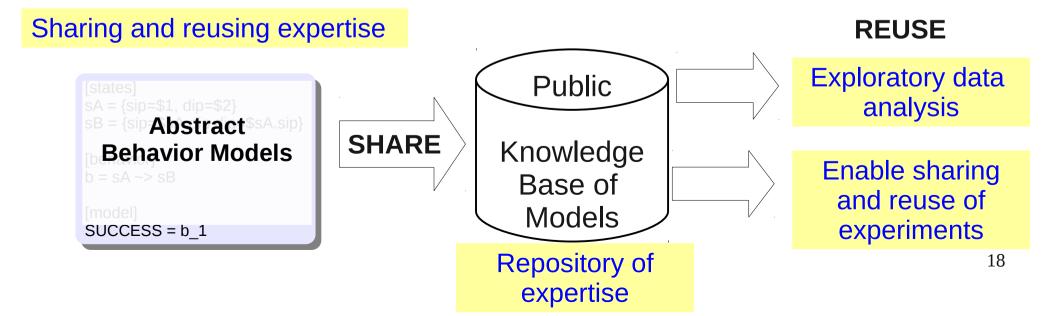
- Extend Modeling Capabities
 - Modeling probabilistic behavior
 - Modeling packet distributions
- Analysis Framework
 - Scalability and performance
 - Reducing the computational complexity of correlations using dependent attributes.

Composing, Sharing and Reusing

Semantic Analysis Framework enables data analysis at higher-levels of abstraction.

Composing models to create higher-level meaning





Thank You!

Our framework will soon be publicly available at <u>http://thirdeye.isi.deterlab.net</u>



Please register on our mailing-list to stay in tune with release and updates