ANALYTICAL CHARACTERIZATION OF INTERNET SECURITY ATTACKS

Doctoral Final Examination
Sarah H. Sellke

PURDUE
UNIVERSITY

OUTLINE

- Overview of My Doctoral Research
- Concealable i.i.d. Timing Channels
- Concealable LRD Timing Channels

INTERNET SECURITY ATTACKS

- Infectious Malware:
  - Worms, Virus
- Concealment Malware:
  - Trojan Horse, rootkit, backdoors
- Malware for Profit:
  - Spyware, Botnet, Keystroke loggers, and dialers
- Data Stealing Malware:
  - Spyware, botnet, key loggers, rootkit, Trojan Horse

2

SUMMARY OF MY DOCTORAL RESEARCH

- Scanning Worms:
  - Uniform and Preference Scanning Worms
  - Stochastic Modeling
  - Automatic Containment of Scanning Worms
- Timing Channels:
  - Capacity Analysis
  - Design of High Rate Network Timing Channels
  - Design of Non-Detectable Network Timing Channels

3

INTERNET SCANNING WORMS (2003 – 2006)

- Stochastic model for the initial growth of infected hosts using branching processes
- Automatic containment schemes
  - Host-based, restricting the total number of scans per host
  - Adaptable to edge-router based containment
- Effectiveness of the containment is shown by both analysis and simulations
- Publications:
  - Conference Proceedings: IEEE Dependable Systems and Networks (DSN), 2005
  - Journal Article: IEEE Transactions on Dependable and Secure Computing (TDSC), 2008

4

TIMING CHANNELS (2006 – 2007)

- Capacity Analysis
  - Found the asymptotically tight upper and lower bounds for timing channels with bounded service time (BSTC)
  - Uniform BSTC has the lowest capacity
  - *Geometric Coding Scheme* achieves capacity
- Publications:
  - Allerton Conference, 2006
  - International Symposium of Information Theory (ISIT), 2007

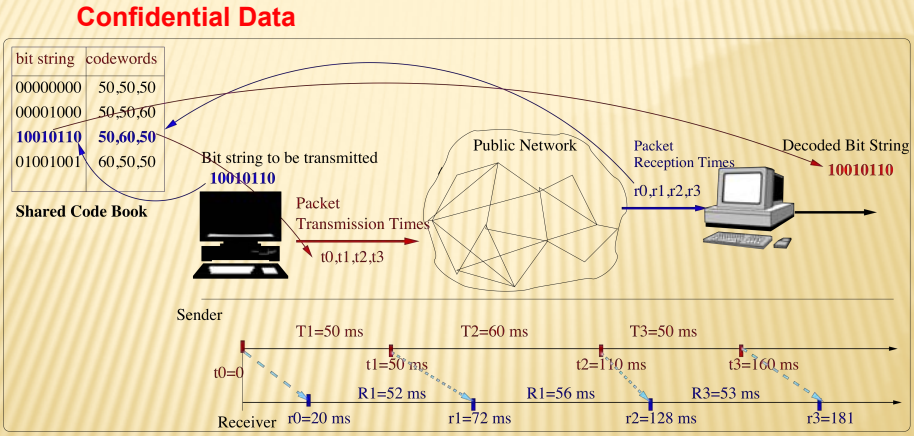
5

NETWORK TIMING CHANNELS (2007 – 09)

- Design and Implementation
  - High Rate Network Timing Channels
    - 5 times current best data rate
  - Non-Detectable Network Timing Channels
    - Independently Distributed Traffic
    - Long Range Dependent Traffic
  - Implemented and Tested on PlanetLab
- Publications:
  - INFOCOM 2009
  - DSN 2010 (submitted)

6

NETWORK COVERT TIMING CHANNELS



OUTLINE

- Overview of My Doctoral Research
- Concealable i.i.d. Timing Channels
- Concealable LRD Timing Channels

8

I.I.D. CONCEALABLE TIMING CHANNEL

- Goal:
  - Immune against current and future detection
- How do we achieved this goal?
  - Mimic the statistical property of i.i.d. normal traffic
  - Computationally indistinguishable from i.i.d. normal traffic

9

CONCEALABLE TIMING CHANNEL

Achieving Design Goals:

- **Mimics statistical property**
- **Computationally indistinguishable from i.i.d. normal traffic**

1: Codeword Look Up.

c(i) –> (x(2i–1), x(2i)) = (15/16, 3/16)

2: Codeword Masking using CSPRNG.

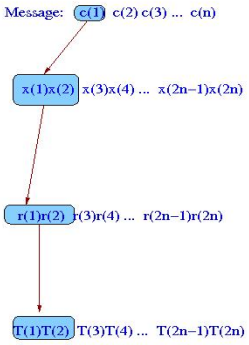
a) CSPRNG –> u(1), u(2), ... u(2n)

b) r(i) = x(i) + u(i) mod 1, for i=1,...,2n

3: Inter–Transmission Time Generation:

$T(i) = F^{-1}(r(i))$

F(x): CDF of a normal traffic



Decoding:

- **Reversal of the above three steps**

10

CONCEALABLE TIMING CHANNEL

- Advantages:
  - Immune from current and future detection
  - Same codebook for different traffic patterns
  - No handshaking necessary
- Disadvantage:
  - Legitimate traffic must be i.i.d.
  - Cannot imitate correlated traffic (HTTP)
    - Easily detectable using second order statistics

11

OUTLINE

- Overview of My Doctoral Research
- Concealable i.i.d. Timing Channels
- Concealable LRD Timing Channels

12

CHALLENGE AND MOTIVATION

- Design network timing channels that can imitate correlated traffic
  - Statistically indistinguishable from real traffic
    - Example: HTTP traffic
  - Evades current best available detection methods
- Usage Scenario:
  - Defeat tight censorship

13

CONCEALABLE TIMING CHANNEL

- Design Goal:
  - Mimic Long Range Dependent Traffic (LRD)
    - HTTP traffic is LRD and non-stationary
  - Desired Marginal Distribution
  - Desired autocorrelation functions
  - Evade current detection methods

14

HOW TO ACHIEVE OUR DESIGN GOAL?

- Modeling of HTTP connection start time
  - Expand an existing model to fit current data
  - CAIDA 2009 trace used
  - FARIMA Model -- Fractional Autoregressive integrated moving average
- Creating Covert Timing Channel
  - Embed covert information in our model
  - Same marginal distribution and autocorrelation

15

RELATED WORK

- W. S. Cleveland *et al*., "IP Packet Generation: Statistical Models for TCP Start Times Based on Connection-Rate Superposition," *in Proceedings of ACM SIGMETRICS, 2000*
  - Analyzed 23 million TCP connections organized into 15 minutes blocks.
  - FARIMA sequence is used to capture LRD
  - the model produces synthetic traffic stochastically similar to that from the actual wire of an Internet link.

16

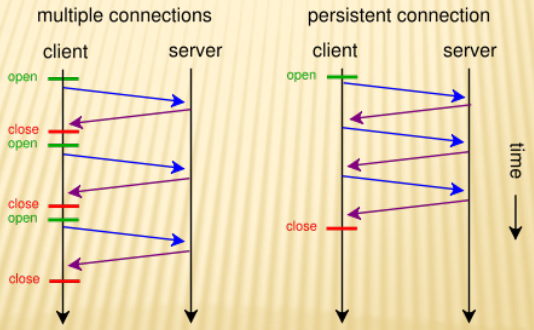
CHALLENGES FOR TIMING CHANNELS

- Stronger Model Requirements:
  - be statistically indistinguishable from the underlying HTTP traffic to avoid detection
  - be able to carry covert information through timing
  - decodable by the receiver

17

PERSISTENT CONNECTIONS IN HTTP/1.1

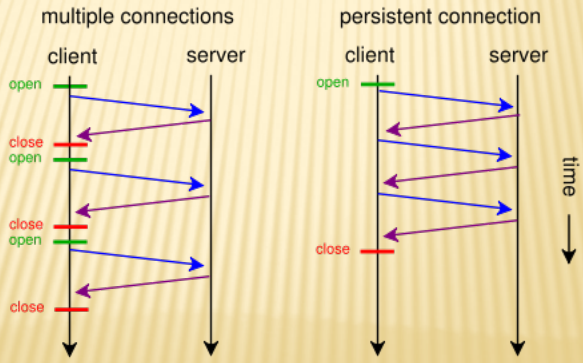
- A single TCP connection is created and reused for multiple HTTP request/response interactions.



18

OUR TIMING CHOICE

- The new TCP connection inter-arrival times is used to carry covert information.



19

EXISTING MODEL

- HTTP new connection inter-arrival times:
  - Non-stationary and LRD
  - the marginal distribution is approximately Weibull

$$F(t) = 1 - e^{-(t/\alpha(r))^{\lambda(r)}}, t \geq 0$$

  - the autocorrelation is modeled by adding white noise to a FARIMA time series.

20

FARIMA MODEL

- Fractional Autoregressive Integrated Moving Average

$$(I - B)^d s_j = \epsilon_j + \epsilon_{j-1}$$

  - ε_j are *i.i.d.* Gaussian Random Variables $(0, \delta_\varepsilon^2)$
  - Fractional values for *d*
  - B is Backward shift operator $Bs_j = s_{j-1}$

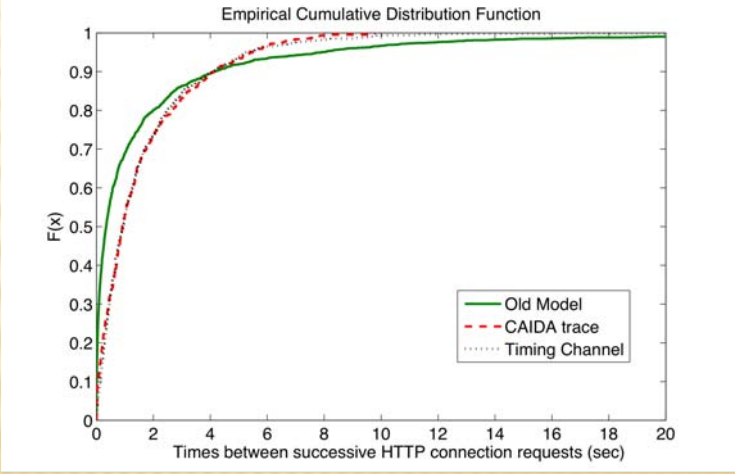
21

LIMITATION OF THE EXISTING MODEL

- Hurst Parameter H:
  - Measures LRD
  - H = d + 0.5
- Existing Model:
  - Developed for Data (1998 – 2000) with Hurst Parameter H=0.75 and Fixed d=0.25
  - Does not fit well for new 2009 data

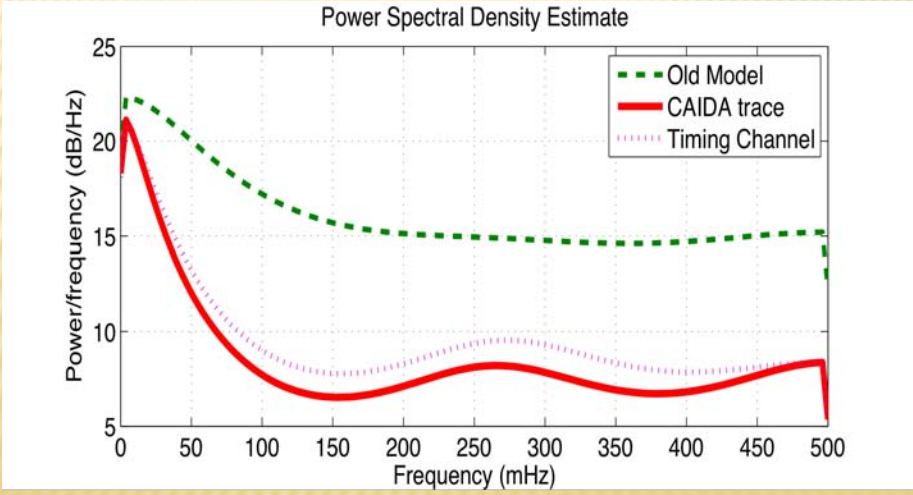
22

LIMITATIONS – MARGINAL DISTRIBUTION



23

LIMITATIONS – SECOND ORDER STATISTICS



24

NEW MODEL

- Expand the existing model:
  - Use the Hurst Parameter estimated from the data in FARIMA model (d=H-0.5)
  - Use the Weibull shape and scale parameters estimated directly from the data to fit the marginal distribution

25

COMPONENTS IN THE NEW MODEL

- FARIMA sequence s_j:

$$(I - B)^d s_j = \epsilon_j + \epsilon_{j-1}$$

- *i.i.d.* Log-Weibull sequence: n_j
- $v_j = s_j + n_j$
- $l_j = g(v_j) = b0 + b1*v_j + b2*v_j^2$
- $T_j = 2^{lj}$

26

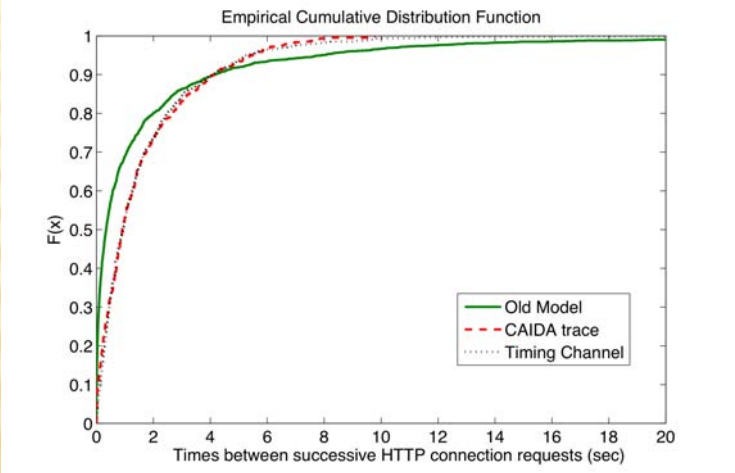
Algorithm 3.1: NEWMODEL$(\alpha, \lambda, H, \rho_1)$

$[1] d = H - 0.5$
$[2] \gamma \leftarrow 0.5772$ //Euler Constant
$[3] \mu_l = \log_2(\alpha) - \gamma \log_2(e)/\lambda,$
$[4] \sigma_l^2 = \pi^2 \log_2^2(e)/6\lambda^2$
$[5] \sigma_s^2 = \sigma_l^2 \rho_1 (2 - d)/(1 + d)$
$[6] \sigma_n^2 = \sigma_l^2 - \sigma_s^2$
$[7] r = 1/(\alpha \Gamma(1 + 1/\lambda))$
$[8] b_0 = 0.7 - e^{-0.7088 - 0.05857r}$
$[9] b_1 = 1 - e^{-1.6301 - 0.06399r}$
$[10] b_2 = -e^{-4.1896 - 0.06254r}$
for $j \leftarrow 1$ **to** n
do
- $[11]$ $s[j] \leftarrow$ FARIMA sequence with variance σ_s^2
- $[12]$ $n[j] \leftarrow i.i.d.$Log-Weibull (μ_l, σ_n^2) sequence
- $[13]$ $v[j] = s[j] + n[j]$
- $[14]$ $l[j] = b_0 + b_1 v[j] + b_2 v^2[j]$
- $[15]$ $t[j] = 2^{l[j]}$

return (t)

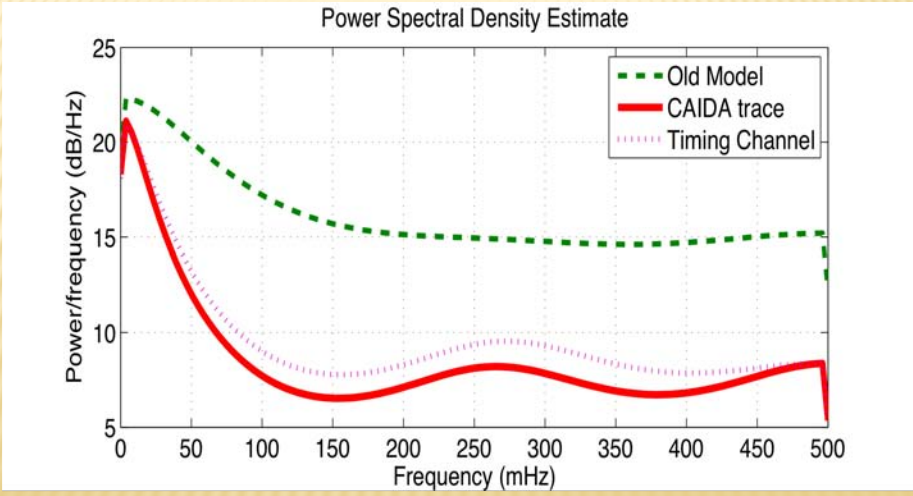
27

LIMITATIONS – MARGINAL DISTRIBUTION



28

LIMITATIONS – SECOND ORDER STATISTICS



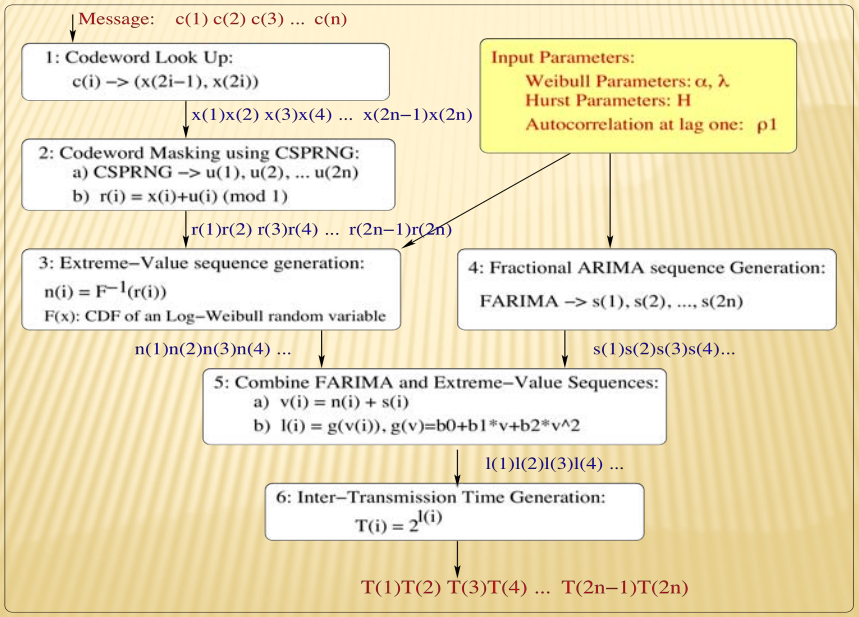
29

DESIGN OF HTTP TIMING CHANNELS

- Synthetic trace from our New Model matches real traffic trace
  - Marginal distribution – Weibull
  - Autocorrelation functions -- LRD
- Embed covert Information in our model
  - Covert information is mapped to inter-arrival times according to our model
  - Retain the statistical properties of the model

30

ENCODER



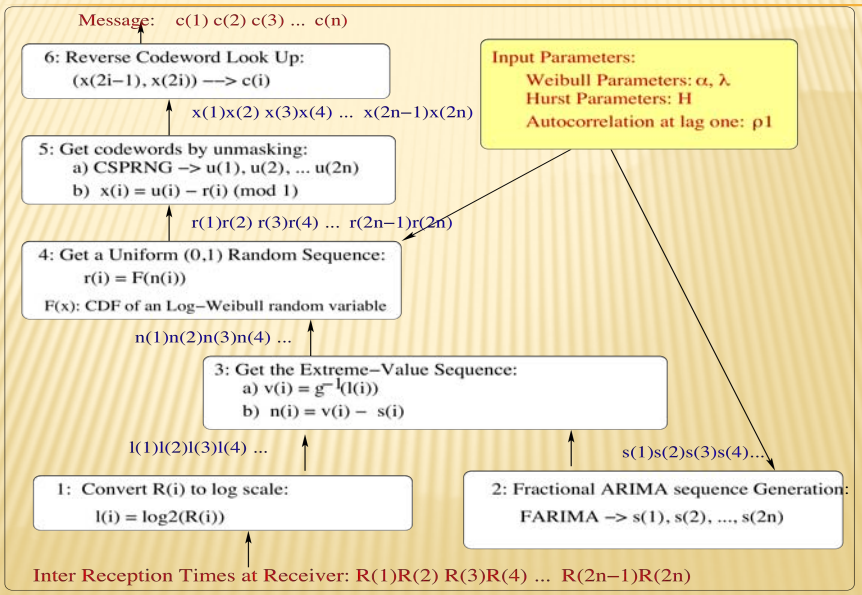
31

SHARING

- Sender and Receiver share:
  - Codebook
    - Character c → (k1/16, k2/16)
  - Traffic Model Parameters:
    - Hurst Parameter H
    - Weibull Parameters: α, λ
    - Autocorrelation at lag 1: ρ1
  - Seeds for CSPRNG and FARIMA sequence

32

DECODER



33

DECODING ERRORS

- Small inter-arrival times cause decoding errors
- Less decoding error when T>100 ms
- Error Correction Method
  - Re-encode the character immediately if the resulting inter-arrival times are small (<100 ms)
  - Require even number of inter-arrival times to represent each character for alignment

34

EXPERIMENTS

- Implementation:
  - Java, Client/Server
  - Thread.sleep(T) for manipulating timing information
- Experiments:
  - PlanetLab Nodes
    - Princeton → Purdue
    - Average RTT 39.5 ms
    - Mimic 8 sets of real data (CAIDA 2009 trace)

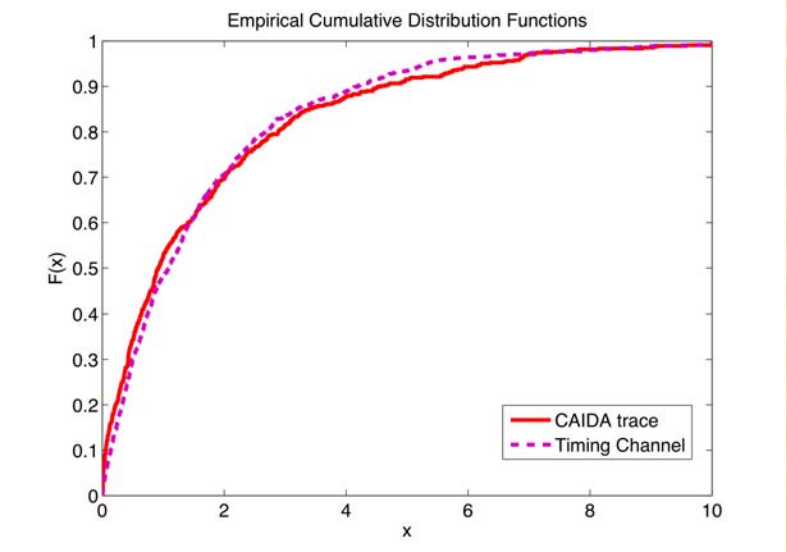
35

INPUT PARAMETERS

SubNet 1 Parameters	data 1	data 2	data 3	data 4
α	1.50	1.34	1.20	1.36
λ	0.77	0.74	0.73	0.76
H	0.61	0.52	0.81	0.65
ρ_1	0.12	0.14	0.23	0.18
r (c/s)	0.57	0.62	0.70	0.64
SubNet 2 Parameters	data 1	data 2	data 3	data 4
α	0.51	0.43	0.44	0.45
λ	0.90	0.85	0.87	0.91
H	0.57	0.56	0.59	0.70
ρ_1	0.08	0.10	0.12	0.09
r (c/s)	1.87	2.15	2.12	2.11

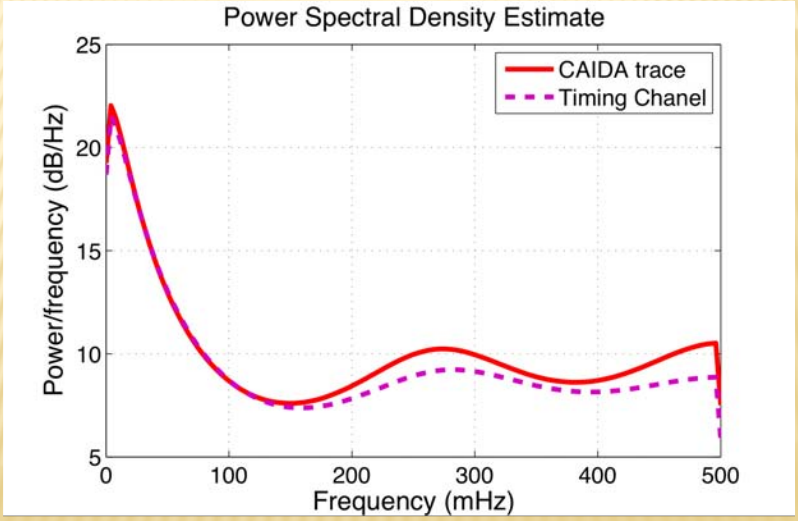
36

MARGINAL DISTRIBUTION



37

SECOND ORDER STATISTICS



38

CURRENT DETECTION METHODS

- *Regularity Test*
  - Traffic is flagged as covert if the variance remains relatively constant
- *Kolmogorov-Smirnov Detection*
  - Traffic is flagged as covert if the maximum distance between the empirical distribution and the desired distribution is too big.
- *Entropy Based Detection*
  - *Equal probable binning w.r.t. the desired distribution*
  - *Entropy detection: marginal distribution*
  - *Conditional Entropy Detection: autocorrelation*

39

KOLMOGOROV-SMIRNOV TEST

SubNet 1	data 1	data 2	data 3	data 4
KS-STAT	0.0657	0.0845	0.0637	0.0516
p-value	0.096	0.012	0.0863	0.2678
detect	no	no	no	no
SubNet 2	data 1	data 2	data 3	data 4
KS-STAT	0.0402	0.0442	0.0466	0.0513
p-value	0.2088	0.1088	0.08	0.043
detect	no	no	no	no

40

ENTROPY DETECTION

traffic type	data 1	data 2	data 3	data 4
training data	6.81	6.80	6.85	6.82
legit	6.56	6.50	6.50	6.51
covert 1	6.85	6.84	6.87	6.89
p-value	0.33	0.36	0.28	0.24
detect?	no	no	no	no

41

CONDITIONAL ENTROPY DETECTION

traffic type	data 1	data 2	data 3	data 4
training data	2.21	2.19	2.17	2.21
legit	1.84	1.81	1.82	1.84
covert 1	2.22	2.23	2.18	2.21
p-value	0.35	0.33	0.44	0.37
detect?	no	no	no	no

42

CONCLUSION

- Designed a covert network timing channel imitating LRD legitimate traffic:
  - can be hidden in the Web traffic, the most observed traffic on Internet today
  - statistically indistinguishable from real traffic
  - evades the best available detection methods.
- Data Rate: 2 – 6 bits/second
- Decoding Error: 3% – 6 %

43

FUTURE RESEARCH DIRECTION

- Develop timing channels for short range dependent (SRD) traffic
- Design a covert timing channel to mimic other commonly used traffics, such as peer-to-peer traffic.

44

Thank You!

45