# Opening Sealed Envelopes

David R. Albrecht

# Big Picture

- DoD ARPAnet: A tragedy of the commons
- “Real Things” on today’s Internet
- Obligatory economic prognostication

# Bandwidth Use

- “Nowadays resource sharing on the Internet is largely a result of what applications, users and operators do at run-time, rather than what the IETF designs into transport protocols at design-time.” -- IETF Working Draft

# C12n Strategies

- Signature-based classification
  - Port-based
  - Payload-based
- Statistical techniques
  - BLINC: BLINd Classification
  - PISA

# Port-Based C12n (I)

- RFC 2324: HTCPCP/1.0 (1 April 1998)
- Nine distinct methods of classifying traffic: (1) port-only, (2) packet header (including port), (3) single-packet signature, (4) single-packet protocol, (5) sig on first KB, (6) first KB protocol, (7) selected flows protocol, (8) all flow protocol, and (9) stateful host tracking

*Moore, AW and Papagiannaki, K. Toward the Accurate Identification of Network Applications. Proceedings of 6th PAM Workshop*

# Port-Based C12n (II)

- (1) port-only
- (2) header (including port)
- (3) single-packet signature
- (4) single-packet protocol
- (5) sig on first KB
- (6) first KB protocol
- (7) selected flows protocol
- (8) all flow protocol, and
- (9) stateful host tracking

| Method | Packets | Bytes |
|---|---|---|
| 1 only | 71.03 | 69.27 |
| 1-2, 9 | 72.05 | 69.38 |
| 1-3, 9 | 72.05 | 69.39 |
| 1-4, 9 | 72.29 | 69.62 |
| 1-5, 9 | 74.23 | 71.48 |
| 1-6, 9 | 80.84 | 78.84 |
| 1-7, 9 | 98.94 | 98.78 |
| All | >99.99 | >99.99 |

*Moore, AW and Papagiannaki, K. Toward the Accurate Identification of Network Applications. Proceedings of 6th PAM Workshop*

# Port-Based C12n (III)

- "A classification technique approaching 100% accuracy proves to be a labor-intensive process that needs to test flow-characteristics against multiple classification criteria in order to gain sufficient confidence in the nature of the causal application."

*Moore, AW and Papagiannaki, K. Toward the Accurate Identification of Network Applications. Proceedings of 6th PAM Workshop*

# VESPA

- In many cases, full protocol parsing is overkill
- Design fast primitives specifically optimized for certain applications?
- Vulnerability-, not exploit-based

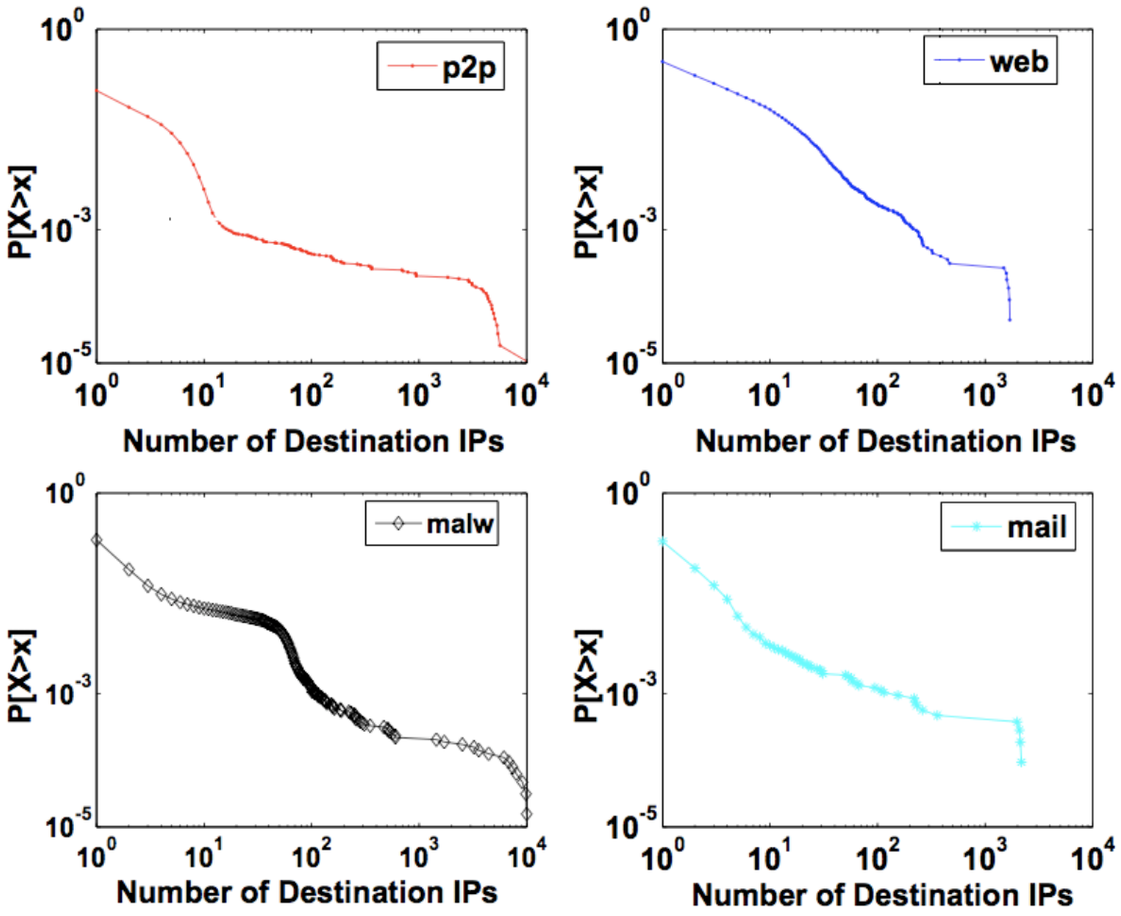
*Schear, N, Albrecht, DR, and Borisov, N. High-speed Matching of Vulnerability Signatures. (Under review)*

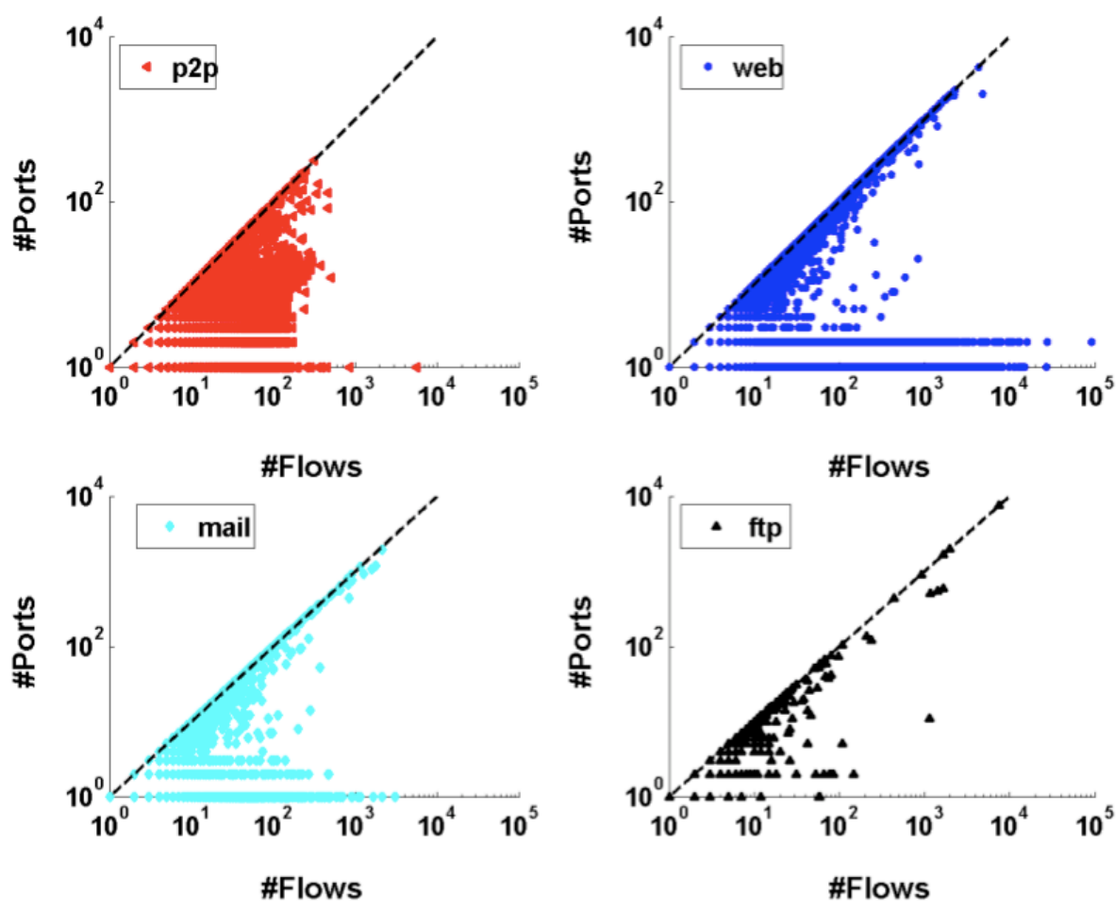
# BLINC

- Blind Classifier
  - Does not look at packet payloads
- Uses social, functional, and application-level characteristics to make application-level judgments

*Karagiannis, T, Papagiannaki, K, and Faloutsos, M. BLINC: Multilevel Traffic Classification in the Dark. SIGCOMM 2005*

# BLINC: Social



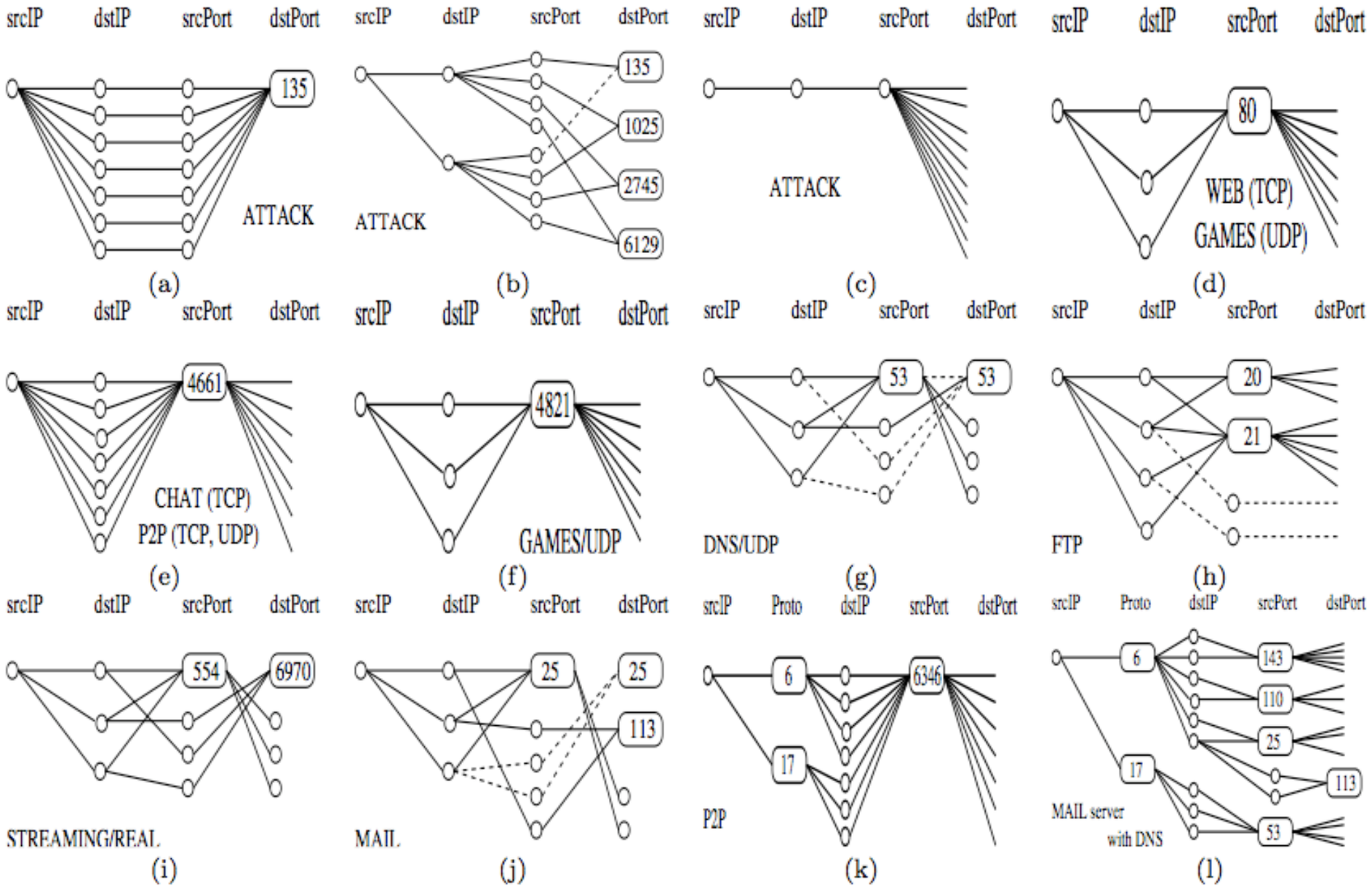

*Karagiannis, T, Papagiannaki, K, and Faloutsos, M. BLINC: Multilevel Traffic Classification in the Dark. SIGCOMM 2005*

# BLINC: Functional

- “For example, let us assume that host A provides a specific service (e.g., web server) and we examine the flows where A appears as a source. Then, A is likely to use a single source port in the vast ma jority of its flows.”

*Karagiannis, T, Papagiannaki, K, and Faloutsos, M. BLINC: Multilevel Traffic Classification in the Dark. SIGCOMM 2005*

# BLINC: Application




*Karagiannis, T, Papagiannaki, K, and Faloutsos, M. BLINC: Multilevel Traffic Classification in the Dark. SIGCOMM 2005*

# BLINC: Redux

- “Our results show that we are able to classify 80%- 90% of the traffic with more than 95% accuracy.”
- “Our first version of BLINC appears…to allow for a real-time implementation… BLINC classified our largest and longest (34-hour) UN2 trace in less than 8 hours"

*Karagiannis, T., Papagiannaki, K., and Faloutsos, M. BLINC: Multilevel Traffic Classification in the Dark. SIGCOMM 2005*

# PISA

- Statistical method for identifying unknown communication protocol
  - Same blindness properties as BLINC
- Uses normed k-dimensional vector space
- Use k-means clustering to find protocol

*Dhamankar, R., and King, R. Protocol Identification via Statistical Analysis. Presented at BlackHat 2007*

# PISA: 10-Space

- “The axes of the PISA space decided by a couple of ‘beer-gut-feelings’ ”

*Dhamankar, R., and King, R. Protocol Identification via Statistical Analysis. Presented at BlackHat 2007*

# PISA: 10-Space

- Shannon entropy of packet data
- client packet size $\mu$ and $\sigma^2$
- server packet size $\mu$ and $\sigma^2$
- client response time $\mu$ and $\sigma^2$
- server response time $\mu$ and $\sigma^2$
- server/client traffic ratio

*Dhamankar, R., and King, R. Protocol Identification via Statistical Analysis. Presented at BlackHat 2007*

# PISA: Results

- Protocol identification
  - Based on 10-Space Euclidean distance
    - Identifies Skype very quickly
- Covert channels
- Again, massive compute overhead

*Dhamankar, R., and King, R. Protocol Identification via Statistical Analysis. Presented at BlackHat 2007*

# Flow Watermarking

- Watermark flows using inter-packet timing
- Spread-spectrums comms techniques to reduce noise

# Errata

- Ability to classify encrypted traffic has much broader implications
  - Many “non-civilians” are not at all aware of SSL et al.’s weaknesses
- Not all p2p is bad! The whole point is efficient content distribution
  - Can help ISPs if used properly

# Conclusions

- ISPs have a strong financial incentive to enforce the “A” of “CIA”
- But, technologies to do this aren’t yet available
- Unfortunately, assuming a cooperative user community is a terrible mistake in today’s Internet climate