Practical Data Hiding in TCP/IP

Kamran Ahsan
and
Deepa Kundur
Bell Canada Junior Chair-holder in Multimedia

Electrical and Computer Engineering
University of Toronto

Agenda

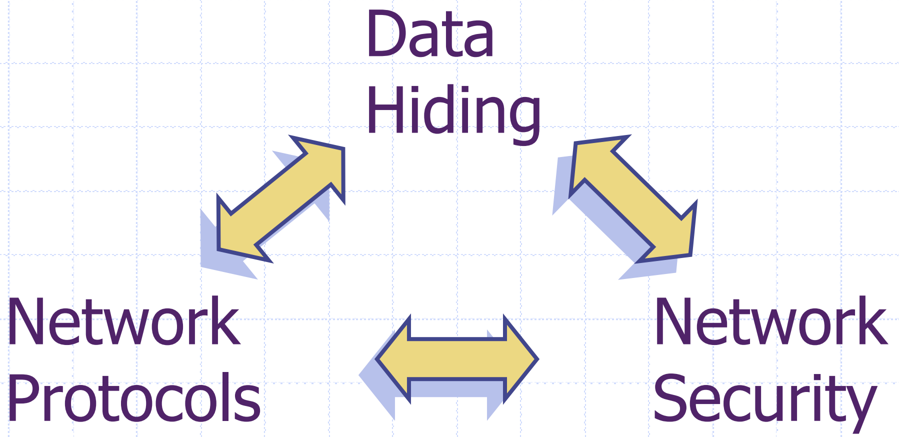
- Introduction
- Problem Formulation
- Previous Work
- Proposed Techniques
- Application Scenarios
- Conclusions

Introduction

- "Open" specifications of the Internet
  - Communications
  - Connectedness
  - Collaboration
- Security in the Internet an afterthought

What is this Paper About?

- Can we identify practical covert channels in TCP/IP?
- How can these channels be used to enhance network processing and security?

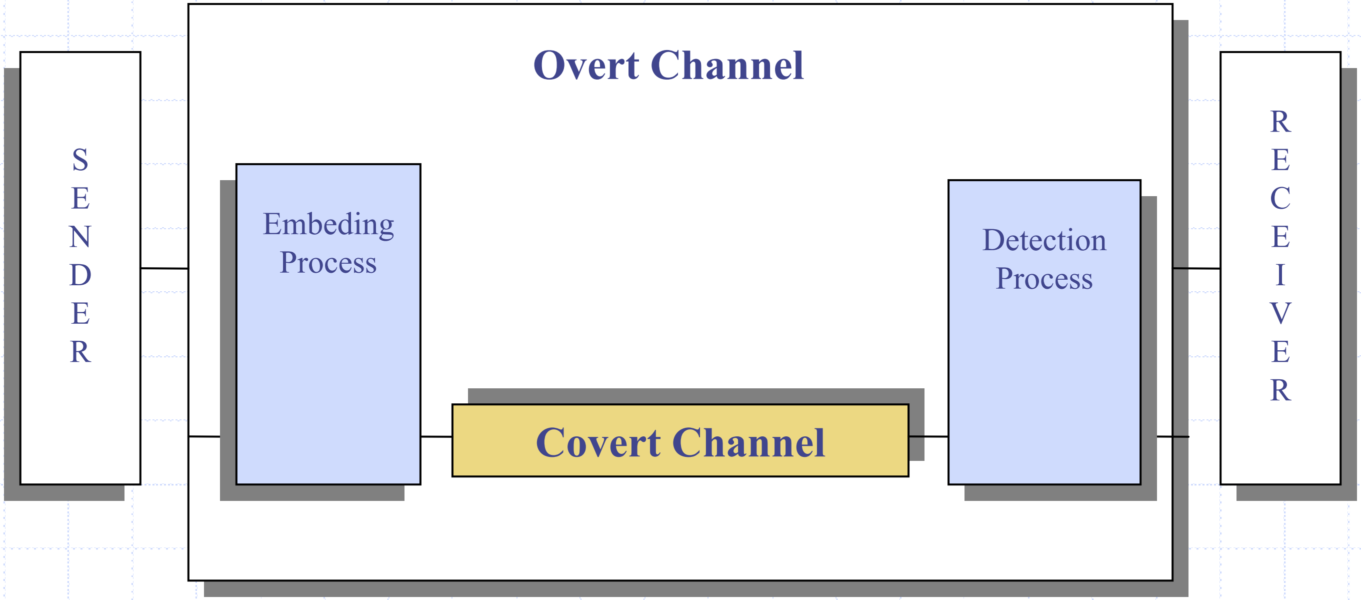


Our Focus

- Covert channels in computer networks
- Data hiding through network packet streams
- Network behavior on packets carrying covert data
- Associated Applications

Covert Channels

- Channel used, but not designed for info transmission
  - Can violate security policy
  - Shared resources, redundancies, multiple interpretations
  - Storage and Timing Channels



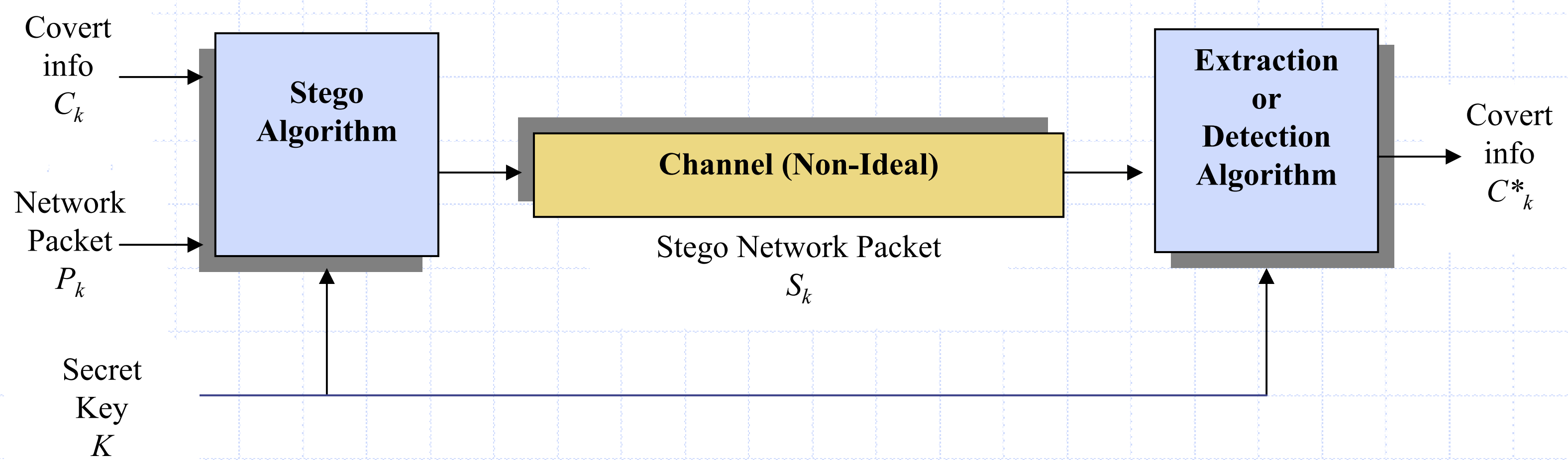
Data Hiding (DH)

- Methodology by which to exploit the presence of covert channels
  - Cover object + Covert Data = Stego-object

- Existing research focused on digital images as cover object

We use network packet streams as the cover object

Framework



- Covert channel piggy-backed on legitimate overt channel
  - Stego Algorithm should not affect overt channel
  - Covert data undetectable by network filters

Previous Work

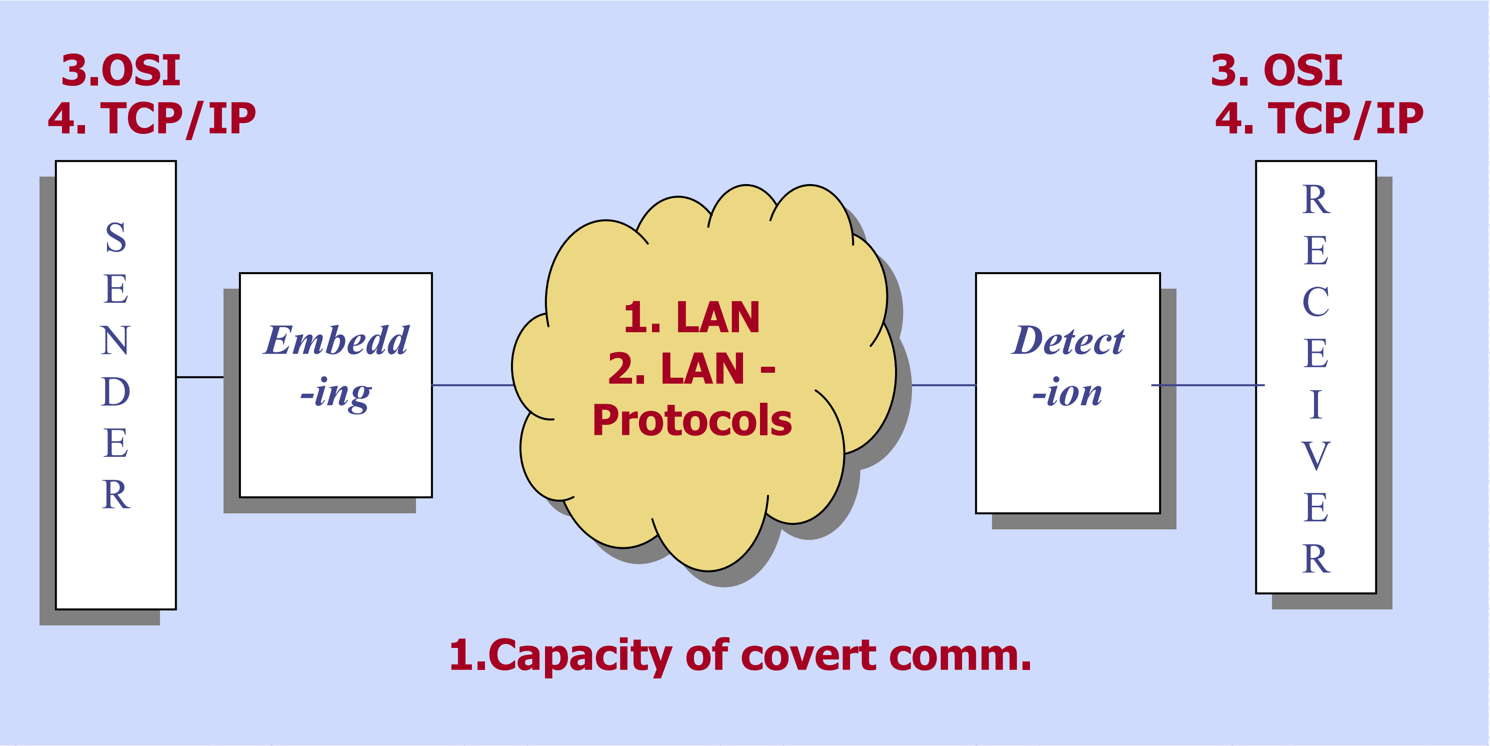
Covert Channel Based

1. Girling (1987): LAN, capacity
2. Wolf (1989): LAN protocols
3. Handel & Sandford (1996): OSI layers
4. Rowland (1997): TCP/IP; proof of the concept

Networks Based

5. Ackermann *et al*. (2000):
   - Weakening of layered concept
   - Additional info. in network packets

The Complete Picture



5. Application Scenarios

Techniques of DH in TCP/IP + (Embedding & Extraction Scenarios) + Application Scenarios + Effects of Covert Communication on Overt Channels + Simulation and Testing

Proposed Algorithms

- Illustrative Examples
  - Packet header manipulation
  - Packet “sorting”

  - Make use of chaotic mixing

Packet Header Manipulation

- IPv4
- Analyzed protocol header
  - Looking for redundancies
  - Multiple interpretations of features and policies
- Develop scenarios wrt network environment

DH Scenario 1

- Multiple interpretation of fragmentation strategy
- Utilize flags field; DF (Do not Fragment) bit

Datagram	16-bit Ident. field	3-bit flag field	13-bit frag. offset	16-bit total length
1	XX…XX	0**1**0	00…00	472

Covertly Communicating ‘**1**’

Datagram	16-bit Ident. field	3-bit flag field	13-bit frag. offset	16-bit total length
2	XX…XX	0**0**0	00…00	472

Covertly Communicating ‘**0**’

DH Scenario 2

- Make use of Sequence Number field
- Must be "unique" for a given source-destination pair

Toral Automorphisms (TAs)

- Chaotic systems
- Watermarking in digital images
- Toral automorphism matrix

$$A = \begin{bmatrix} 1 & 1 \\ k & k+1 \end{bmatrix}$$

maps all points on a lattice

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ k & k+1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} \pmod{K}$$

periodic; recurrence time, *P*

Toral Automorphisms(2)

- Generation of sequence numbers:
  - **Main key** = Size of the Lattice; K
  - **Sub key** = Parameter of TA matrix; k
  - **Third key** = No. of TA applications

TAs provides structured scrambling
and enables Alice and Bob to communicate
covertly

DH Scenario 2

- Alice's End
  1. Selection of keys
  2. Formation of a look-up table; sorted sequence matched with alphabet
  3. Conversion to binary
  4. Appending randomly generated 8 bits to form 16-bit Identification field
- Bob's End
  - Generation of the look-up table
  - Deciphering of the Identification field thereafter

DH Scenario 2

- Generation of Identifier by chaotic mixing

k=1,
K=26,
third key=8

Sr. #	Alphabets	Seq.for 8th iter.	Binary Equ.	Encoded in 4-bit	Ident.Field
1	A	1	0000 00001	0 1	0 1 X X
2	B	14	0000 1110	0 E	0 E X X
3	C	9	0000 1001	0 9	0 9 X X
4	D	22	0001 0110	16	1 6 X X
5	E	4	0000 0100	0 4	0 4 X X
6	F	17	0001 0001	1 1	1 1 X X
7	G	25	0001 1001	1 9	1 9 X X
8	H	12	0000 1100	0 C	0 C X X
24	X	24	0001 1000	1 8	1 8 X X
25	Y	6	0000 0110	0 6	0 6 X X
26	Z	19	0001 0011	1 3	1 3 X X

Potential Applications

- Enhanced filtering criteria in firewalls
- Security tied to the content – client-server architecture
- Content delivery networks

Data Hiding by Packet Sorting

- Sorting: 'n' objects can store $\log_2(n!)$ bits
- Packet "sorting" / "resorting" at network layer
  - Reference = Sequence number field of IPSec
  - No major modification in header fields
  - Sorting: chaotic mixing
  - Resorting: best sequence estimation

Sorting / Resorting Process

- Two keys:
  - Main key = K
  - Sub key = k
- Covert data:

  third key = seq. no.

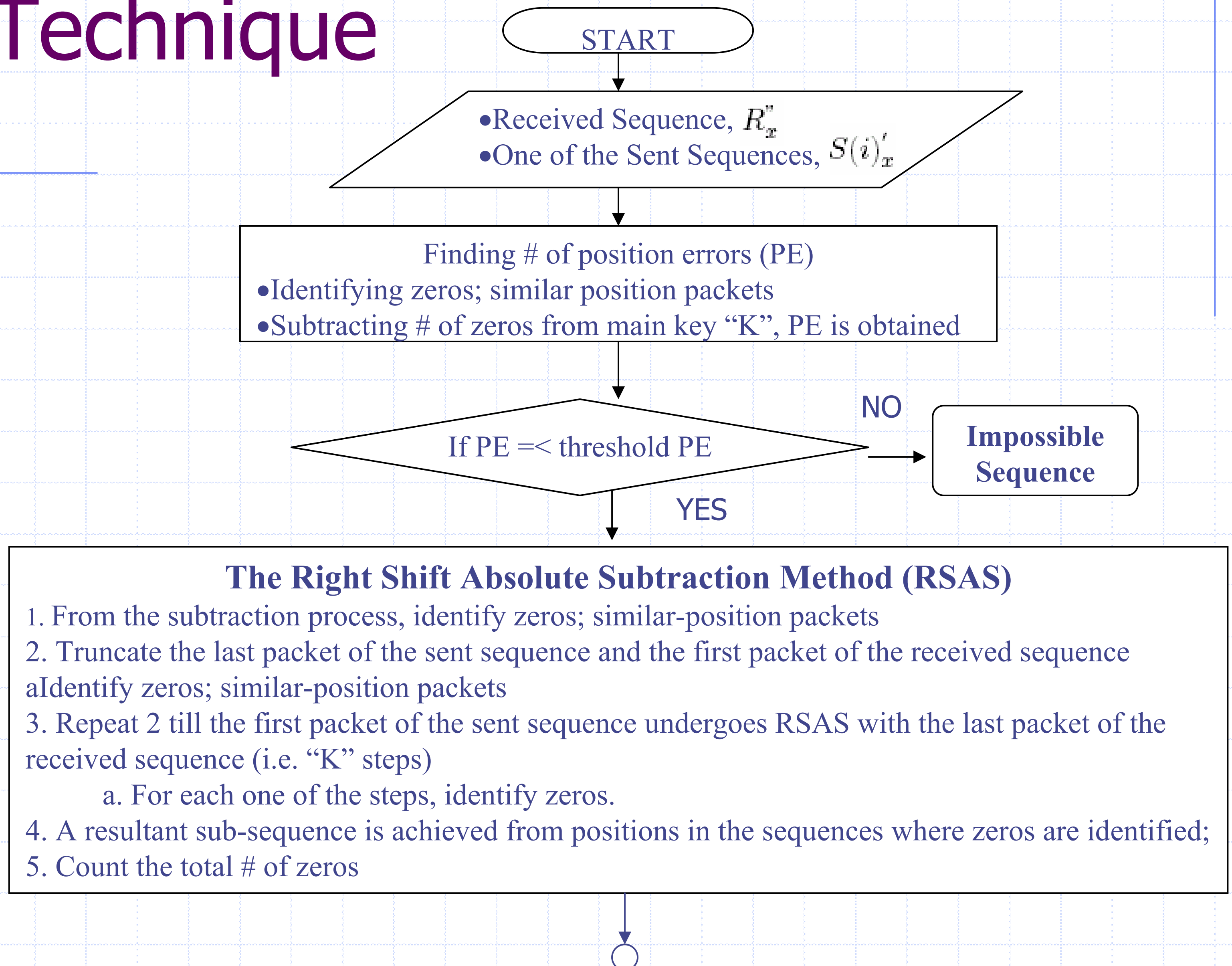
- From
  - Main key = K
  - Sub key = k

Bob *estimates* the third key, the covert message

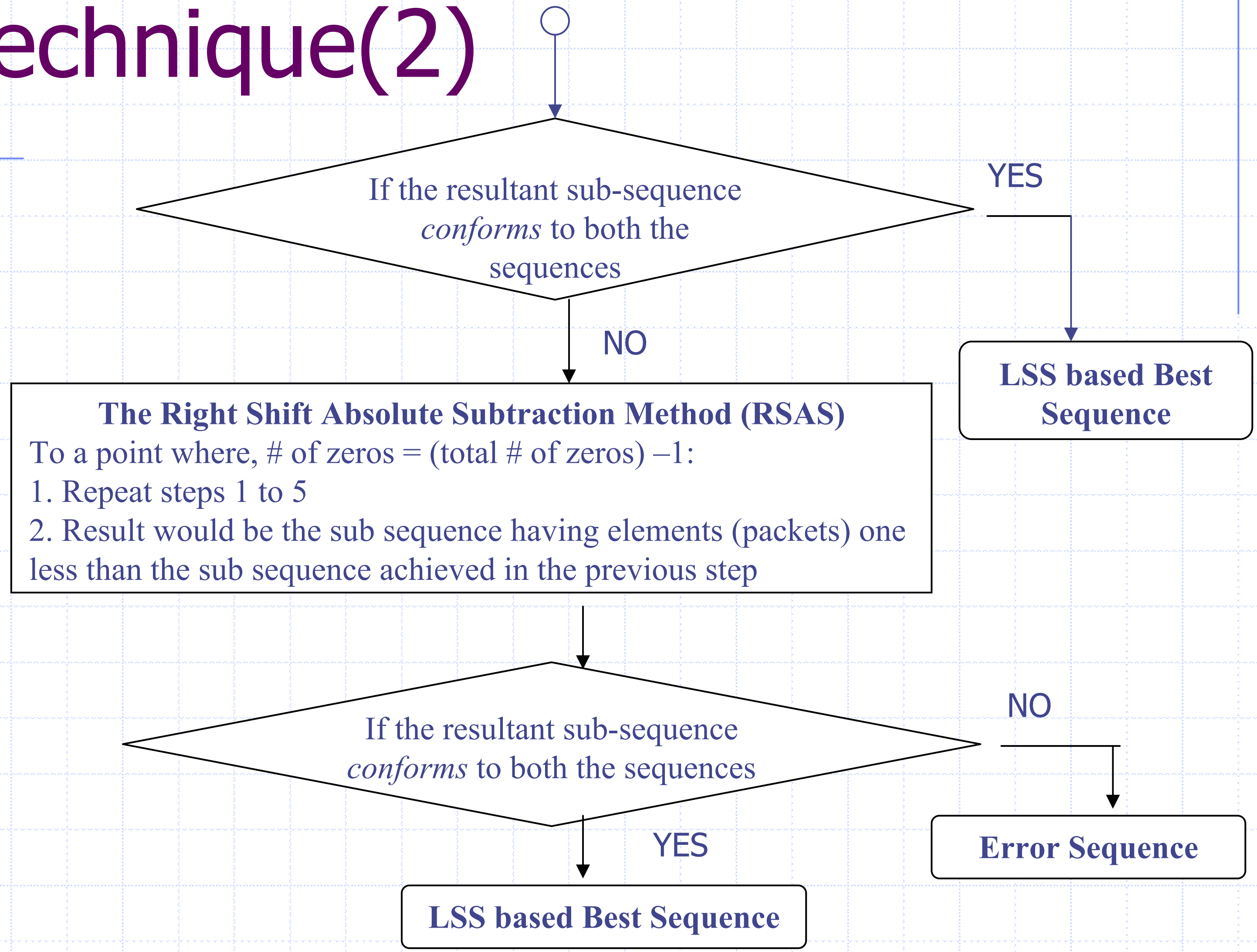
Best Sequence Estimation

- Out of order delivery by the Internet layer
  - Out of orderedness is prevalent and asymmetric
  - Introduction of packet position errors
  - Small scale reordering ; Paxson and Mogul findings
- Longest Subsequence (LSS) Method

The Technique

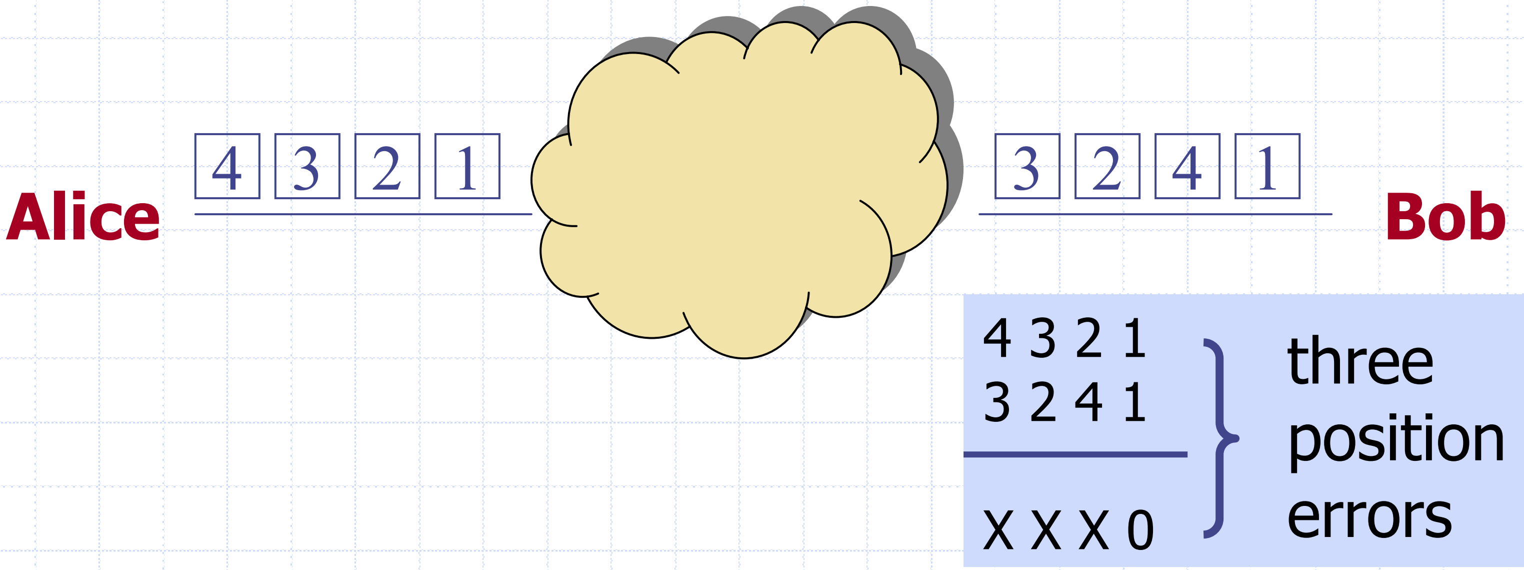


The Technique(2)



Simulation and Testing

- Process is simulated for K=4,…,8 and k=1
- A practical communication network; introducing 3 to 6 position errors in packet sequence.



Position Error (PE) Scenarios

- Small scale position errors (Paxson & Mogul)
- For specific packet sequence: Consider
  - All position errors
  - All permutations equally likely
- Evident sequences or/and LSS based best estimate sequences are highly desirable

Analysis - PE Scenarios

Which sent sequence is most likely to be mapped at Bob's end?

Received Sequence Category	Seq.1 S(1)	Seq.2 S(2)	Seq.3 S(3)	Seq.4 S(4)	Total
Impossible	-	-	-	-	541
Error	-	-	-	-	15
Evident	33	36	40	34	143
Best Estimate	5	5	4	7	21
Total	38	41	44	41	720

Main key= 6(imprvd.); Sent seq.= 4; Network behavior: 3 PE or less

Usage Scenarios

- Packet Sorting/Resorting
  - Preliminary authentication in IPSec
  - Enhanced anti-traffic analysis
  - Enhanced security mechanisms for IPSec protocols

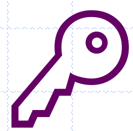
Conclusion

- Covert channels in networks – a step forward
- Network processing and security can be reinforced by integrating steganography with existing security architecture
  - Packet header manipulation – network processing and security services
  - Packet sorting – network security mechanisms

Q&A

$$A = \begin{bmatrix} 1 & 1 \\ k & k+1 \end{bmatrix}$$

IPSec



3	2	4	1

4	3	2	1

<table>
<tr><td>4-bit Ver.
0100</td><td>4-bit IHL
0101</td><td>8-bit TOS
XXXXXXUU</td><td colspan="2">16-bit Tot. Len.
XXXXXXXXXXXXXXXX</td></tr>
<tr><td colspan="3">16-bit Ident.
0000 0100 RRRRRRRR</td><td>3-bit flags
XXX</td><td>13-bit Frag.Off.
XXXXXXXXXXXXX</td></tr>
<tr><td colspan="2">8-bit TTL
XXXXXXXX</td><td>8-bit Protocol
XXXXXXXX</td><td colspan="2">16-bit Checksum
XXXXXXXXXXXXXXXX</td></tr>
<tr><td colspan="5">32-bit Source Address
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX</td></tr>
<tr><td colspan="5">32-bit Destination Address
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX</td></tr>
</table>

IPv4