

DATA MINING FOR NETWORK INTRUSION DETECTION

Vipin Kumar

Army High Performance Computing Research Center

Department of Computer Science

University of Minnesota

<http://www.cs.umn.edu/~kumar>

Collaborators: Paul Dokas, Eric Eilertson, Levent Ertoz, Yongdae Kim,
Aleksandar Lazarevic, Jaideep Srivastava, Pang-Ning Tan,
Zhi-li Zhang

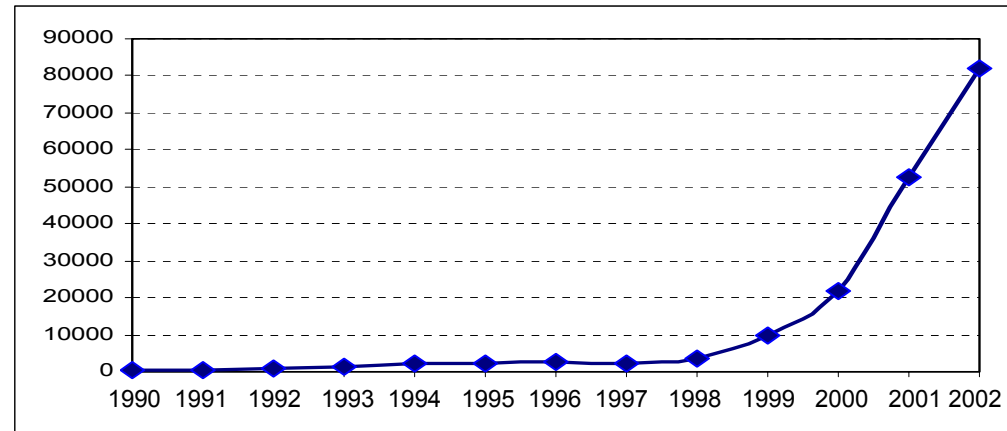
Research supported by AHPCRC/ARL



Introduction

- ◆ As the cost of information processing and Internet accessibility falls, more organizations are becoming increasingly vulnerable to potential cyber threats such as network intrusions

Incidents Reported to Computer Emergency Response Team/Coordination Center (CERT/CC)

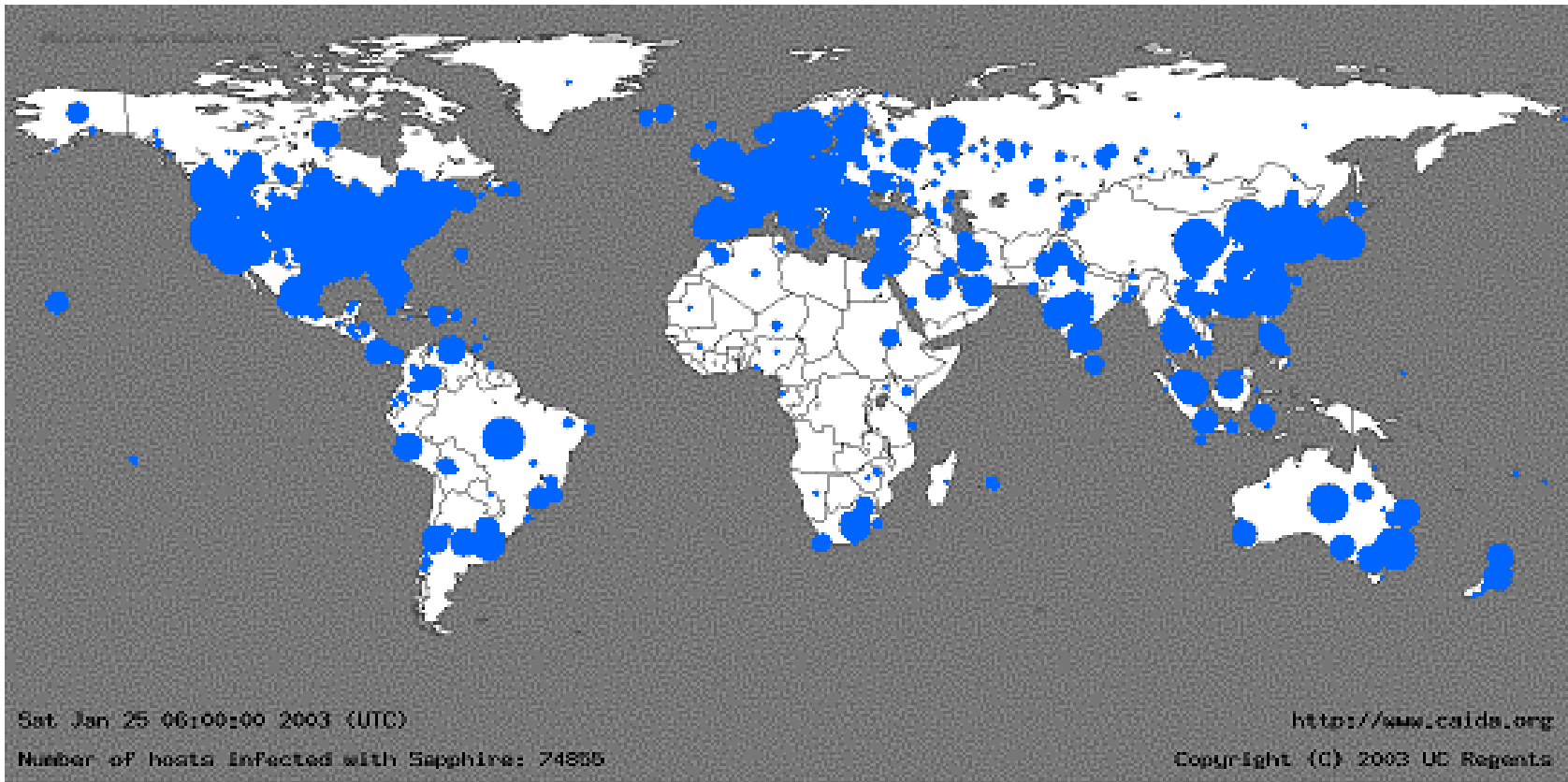


- ◆ There is an increasing awareness around the world about the danger that is coming from cyber world through different cyber attacks



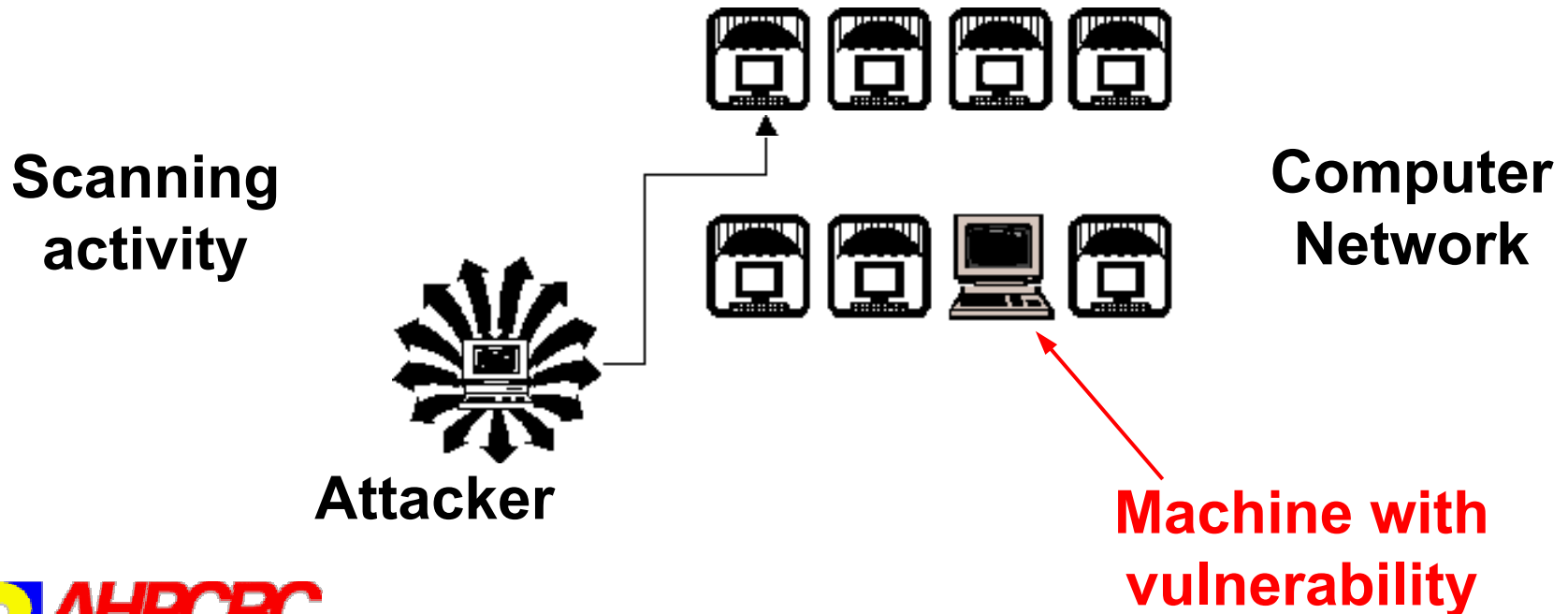
The Spread of the Sapphire/Slammer Worm

- The geographic spread of Sapphire/Slammer Worm 30 minutes after release



What are Intrusions?

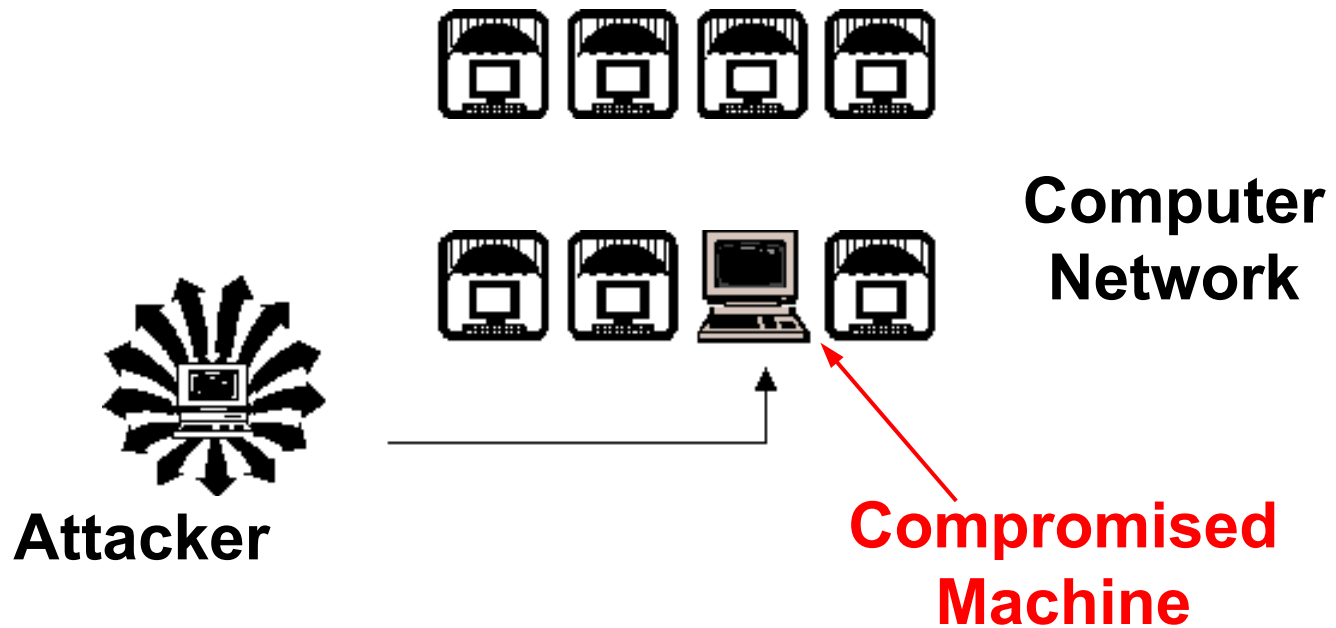
- ◆ Intrusions are actions that attempt to bypass security mechanisms of computer systems. They are caused by:
 - ◆ Attackers accessing the system from Internet
 - ◆ Insider attackers - authorized users attempting to gain and misuse non-authorized privileges
- ◆ Typical intrusion scenario



What are Intrusions?

- ◆ Intrusions are actions that attempt to bypass security mechanisms of computer systems. They are caused by:
 - ◆ Attackers accessing the system from Internet
 - ◆ Insider attackers - authorized users attempting to gain and misuse non-authorized privileges

- ◆ Typical intrusion scenario



Why We Need Intrusion Detection Systems

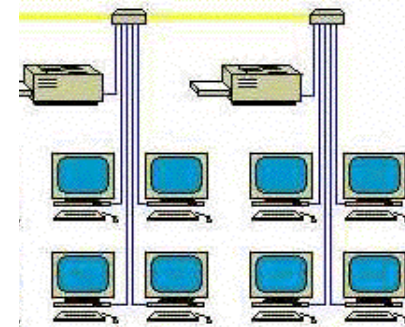
- ◆ **Security mechanisms always have inevitable vulnerabilities**
- ◆ **Current firewalls are not sufficient to ensure security in computer networks**
 - ◆ **“Security holes” caused by allowances made to users/programmers/administrators**
 - ◆ **Insider attacks**
 - ◆ **Multiple levels of data confidentiality needs multi-layer protection in firewalls**



Intrusion Detection

◆ Intrusion Detection System

- ◆ combination of software and hardware that attempts to perform intrusion detection
- ◆ raises the alarm when possible intrusion happens



◆ Traditional intrusion detection system IDS tools (e.g. SNORT) are based on signatures of **known attacks**

- ◆ Example of SNORT rule (**MS-SQL “Slammer” worm**)

```
any -> udp port 1434 (content:"|81 F1 03 01 04 9B 81 F1 01|";  
content:"sock"; content:"send")
```



www.snort.org

◆ Limitations

- ◆ Signature database has to be manually revised for each new type of discovered intrusion
- ◆ **They cannot detect emerging cyber threats**
- ◆ Substantial latency in deployment of newly created signatures across the computer system

- Data mining can alleviate these limitations

Data Mining for Intrusion Detection

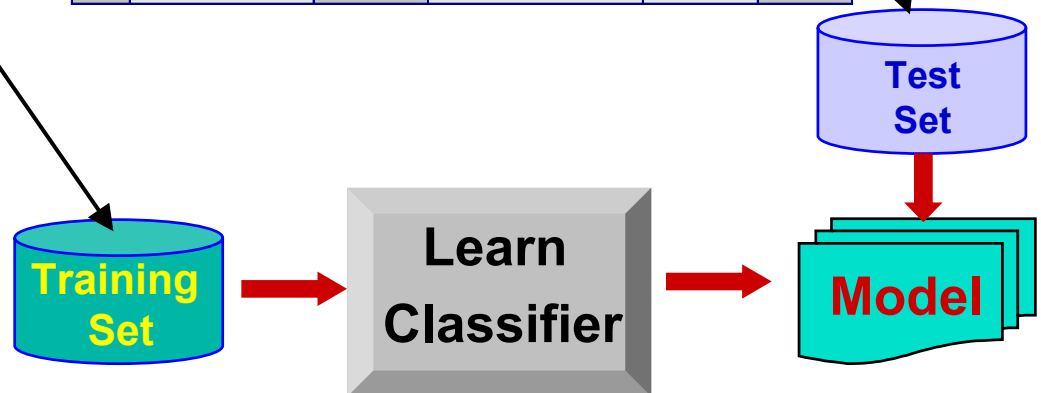
Misuse detection – Building Predictive Models

- Predictive models are built from labeled data sets (instances are labeled as “normal” or “intrusive”)
- These models can be more sophisticated and precise than manually created signatures
- Unable to detect attacks whose instances have not yet been observed

categorical temporal categorical continuous class

Tid	SrcIP	Start time	Dest Port	Number of bytes	Attack
1	206.135.38.95	11:07:20	160.94.179.223	192	No
2	206.163.37.95	11:13:56	160.94.179.219	195	No
3	206.163.37.95	11:14:29	160.94.179.217	180	No
4	206.163.37.95	11:14:30	160.94.179.255	199	No
5	206.163.37.95	11:14:32	160.94.179.254	19	Yes
6	206.163.37.95	11:14:35	160.94.179.253	177	No
7	206.163.37.95	11:14:36	160.94.179.252	172	No
8	206.163.37.95	11:14:38	160.94.179.251	285	Yes
9	206.163.37.95	11:14:41	160.94.179.250	195	No
10	206.163.37.95	11:14:44	160.94.179.249	163	Yes

Tid	SrcIP	Start time	Dest Port	Number of bytes	Attack
1	206.163.37.81	11:17:51	160.94.179.208	150	?
2	206.163.37.99	11:18:10	160.94.179.235	208	?
3	206.163.37.55	11:34:35	160.94.179.221	195	?
4	206.163.37.37	11:41:37	160.94.179.253	199	?
5	206.163.37.41	11:55:19	160.94.179.244	181	?



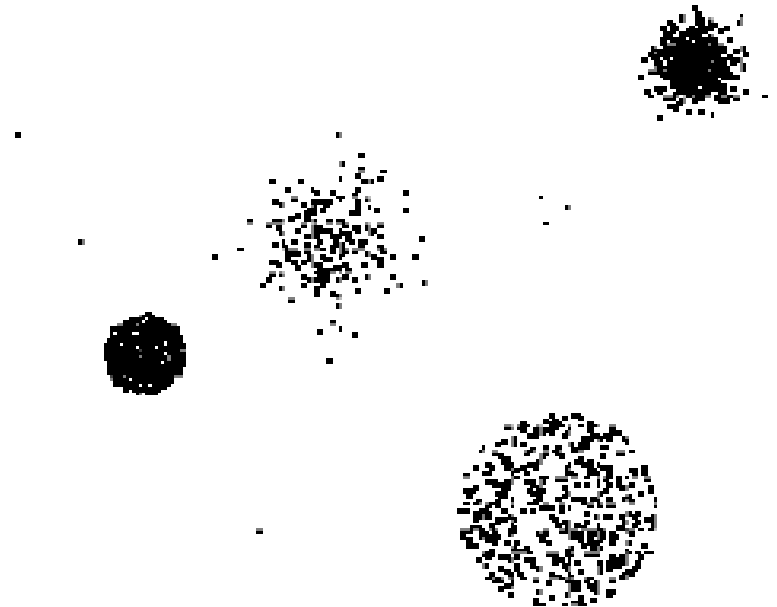
Data Mining for Intrusion Detection

▪ *Anomaly detection*

- Identifies anomalies as deviations from “normal” behavior
- Potential for high false alarm rate - previously unseen (yet legitimate) system behaviors may also be recognized as anomalies

▪ *Recent research*

- Stolfo, Lee, et al; Barbara, Jajodia, et al; James; Lippman et al; Bridges et al; etc.



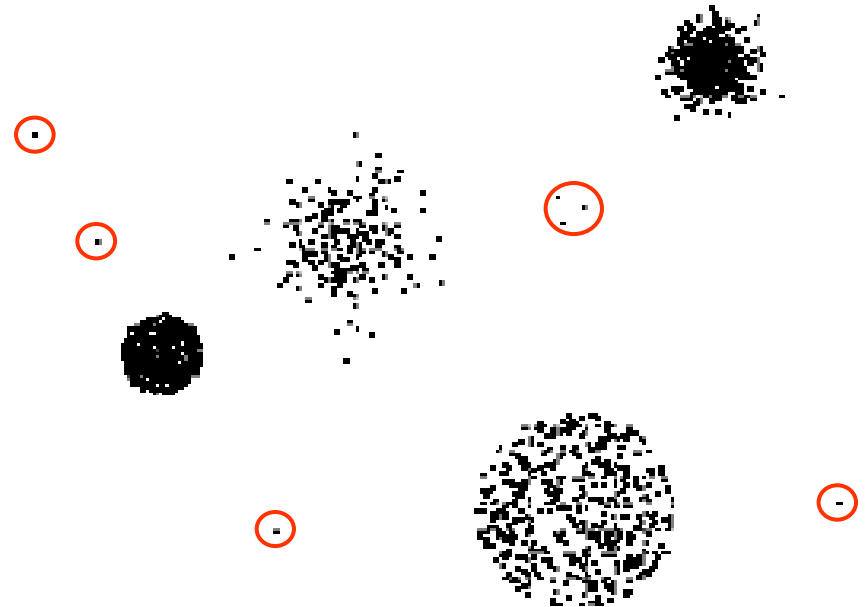
Data Mining for Intrusion Detection

▪ *Anomaly detection*

- Identifies anomalies as deviations from “normal” behavior
- Potential for high false alarm rate - previously unseen (yet legitimate) system behaviors may also be recognized as anomalies

▪ *Recent research*

- Stolfo, Lee, et al; Barbara, Jajodia, et al; James; Lippman et al; Bridges et al; etc.



Data Mining for Intrusion Detection

Summarization of attacks using association rules

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk



Rules Discovered:

{Milk} --> {Coke}

{Diaper, Milk} --> {Beer}

Tid	SrcIP	Start time	Dest IP	Dest Port	Number of bytes
1	206.163.X.95	11:07:20	160.94.179.223	139	192
2	206.163.X.95	11:13:56	160.94.179.219	139	195
3	206.163.X.95	11:14:29	160.94.179.217	139	180
4	206.163.X.95	11:14:30	160.94.179.255	139	199
5	206.163.X.95	11:14:32	160.94.179.254	139	186
6	206.163.X.95	11:14:35	160.94.179.253	139	177
7	206.163.X.95	11:14:36	160.94.179.252	139	172
8	206.163.X.95	11:14:38	160.94.179.251	139	192
9	206.163.X.95	11:14:41	160.94.179.250	139	195
10	206.163.X.95	11:14:44	160.94.179.249	139	163



Rules Discovered:

**{Src IP = 206.163.X.95,
Dest Port = 139,
Bytes ∈ [150, 200]} --> {ATTACK}**

Key Technical Challenges

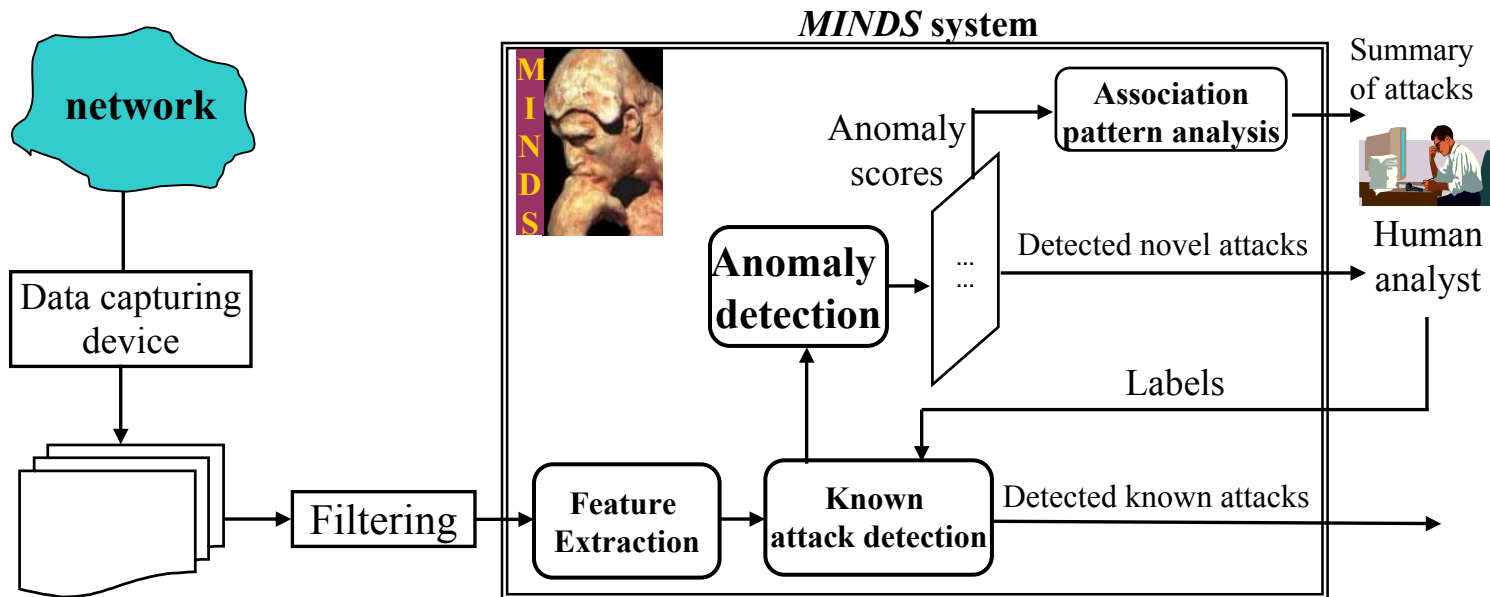
- ◆ **Large data size**
 - ◆ Millions of network connections are common for commercial network sites, ...
- ◆ **High dimensionality**
 - ◆ Hundreds of dimensions are possible
- ◆ **Temporal nature of the data**
 - ◆ Data points close in time - highly correlated
- ◆ **Skewed class distribution**
 - ◆ Interesting events are very rare \Rightarrow looking for the “needle in a haystack”
- ◆ **Data Preprocessing**
 - ◆ Converting network traffic into data
- ◆ **High Performance Computing (HPC) is critical for on-line analysis and scalability to very large data sets**



“Mining needle in a haystack.
So much hay and so little time”

The MINDS Project

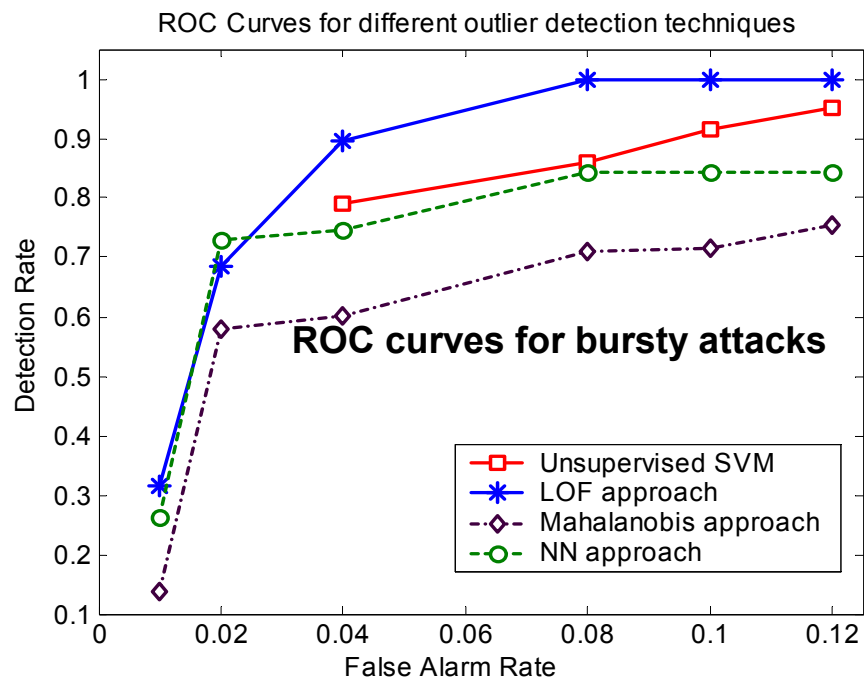
◆ MINDS – MINnesota INtrusion Detection System



Feature construction

- **Three groups of features**
 - ◆ **Basic features of individual TCP connections**
 - source & destination IP/port, protocol, number of bytes, **duration**, **number of packets** (used in SNORT only in stream builder module)
 - ◆ **Time based features**
 - For the same source (destination) IP address, number of unique destination (source) IP addresses inside the network *in last T seconds*
 - Number of connections from source (destination) IP to the same destination (source) port *in last T seconds*
 - ◆ **Connection based features**
 - For the same source (destination) IP address, number of unique destination (source) IP addresses inside the network *in last N connections*
 - Number of connections from source (destination) IP to the same destination (source) port *in last N connections*

Anomaly Detection on DARPA'98 Intrusion Detection Benchmark Data

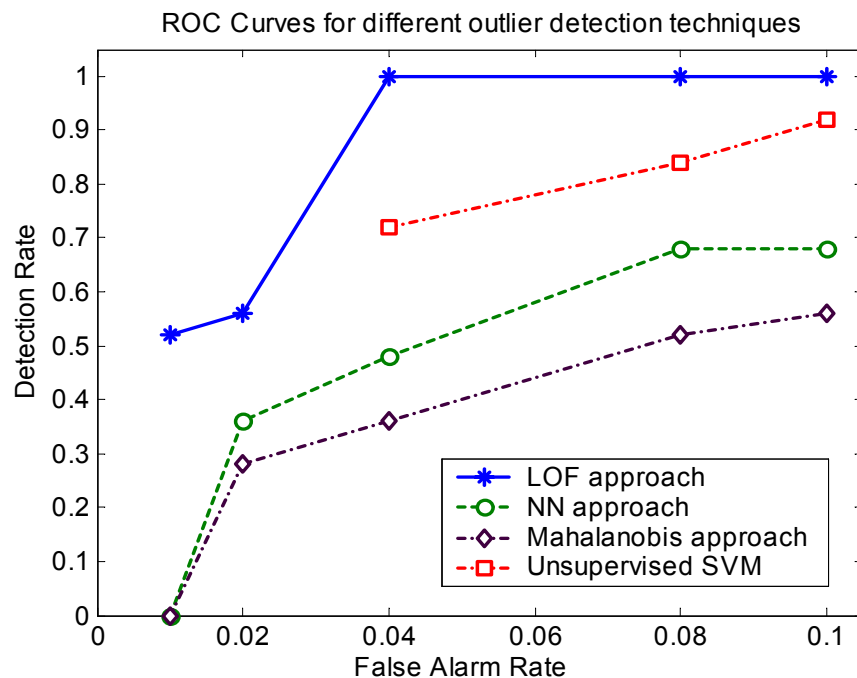


LOF approach is consistently better than other approaches

Unsupervised SVMs are good but only for high false alarm (FA) rate

NN approach is comparable to LOF for low FA rates, but detection rate decrease for high FA

Mahalanobis-distance approach – poor due to multimodal normal behavior



ROC curves for **single-connection** attacks

LOF approach is superior to other outlier detection schemes

Majority of single connection attacks are probably located close to the dense regions of the normal data



Anomaly Detection on Real Network Data

- During the past few months various intrusive/suspicious activities were detected at the AHPCCRC and at the U of Minnesota using *MINDS*
- Many of these could not be detected using state-of-the-art tool like SNORT
- Anomalies/attacks picked by *MINDS*
 - ◆ Scanning activities
 - ◆ Non-standard behavior
 - Policy violations
 - Worms

Detection of Scans on Real Network Data

- August 13, 2002

- ◆ **Detected scanning for Microsoft DS service on port 445/TCP (Ranked #1)**

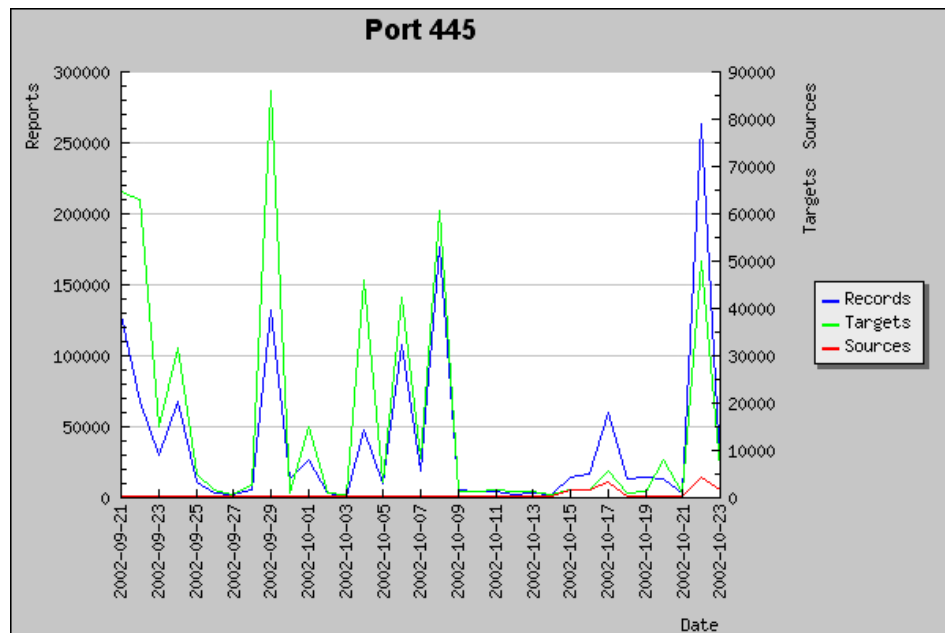
- Reported by CERT as recent DoS attacks that needs further analysis

- (CERT August 9, 2002)

- Undetected by SNORT since the scanning was non-sequential (very slow)

- A rule added to SNORT later in September

Number of scanning activities on Microsoft DS service on port 445/TCP reported in the World
(Source www.incidents.org)



- August 13, 2002

- ◆ **Detected scanning for Oracle server (Ranked #2)**

- ◆ Reported by CERT, June 13, 2002

- ◆ First detection of this attack type by our University

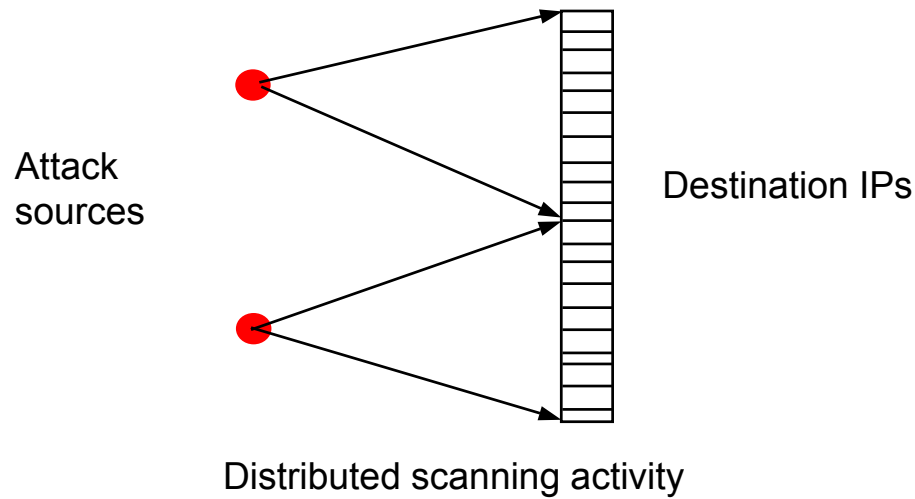
- ◆ Undetected by SNORT because the scanning was hidden within another Web scanning



Detection of Scans on Real Network Data

◆ October 10, 200

- **Detected a distributed windows networking scan from multiple source locations (Ranked #1)**
- **Similar distributed scan from 100 machines scattered around the World happened at University of Auckland, New Zealand, on August 8, 2002 and it was reported by CERT, Insecure.org and other security organizations**



Detection of Policy Violations on Real Network Data

- August 8, 2002

- ◆ **Identified machine that was running Microsoft PPTP VPN server on non-standard ports, which is a policy violation (Ranked #1)**
 - ◆ Undetected by SNORT since the collected GRE traffic was part of the normal traffic

- August 10 2002, October 30, 2002

- ◆ **Identified compromised machines that were running FTP servers on non-standard ports, which is a policy violation (Ranked #1)**
 - Anomaly detection identified this due to huge file transfer on a non-standard port
 - Undetectable by SNORT due to the fact there are no signatures for these activities
 - **Example of anomalous behavior following a successful Trojan horse attack**

Detection of Policy Violations on Real Network Data

◆ January 27, 2003

- **Detected odd, not routable RFC1918 traffic coming from the Internet**
 - ◆ RFC1918 (Request for Comments) serves as Address Allocation for Private Internets. RFC1918 blocks are segments of IP address space reserved by IANA (Internet Assigned Numbers Authority) for use within an organization
 - ◆ DNS records for RFC1918 addresses are legitimate only within the network on which a host with RFC1918 address resides
 - ◆ RFC1918 addresses are not globally routed and they should not appear on the public Internet

◆ February 6, 2003

- **The IP address 128.101.X.0 (not a real computer, but a network itself) has been targeted with IP Protocol 0 traffic from Korea (61.84.X.97)**
 - ◆ This is “exceedingly” bad as IP Protocol 0 is not legitimate.

Detection of Policy Violations on Real Network Data

◆ February 6, 2003

- **Detected a computer on the network apparently communicating with a computer in California over a VPN.**
 - ◆ **Worst case:** This is a covert channel by which someone might be gaining access to the University network in an unauthorized way.
 - ◆ **Best case:** This is someone at the University creating unauthorized tunnels between the University and some other network, which is not allowed.

◆ February 7, 2003

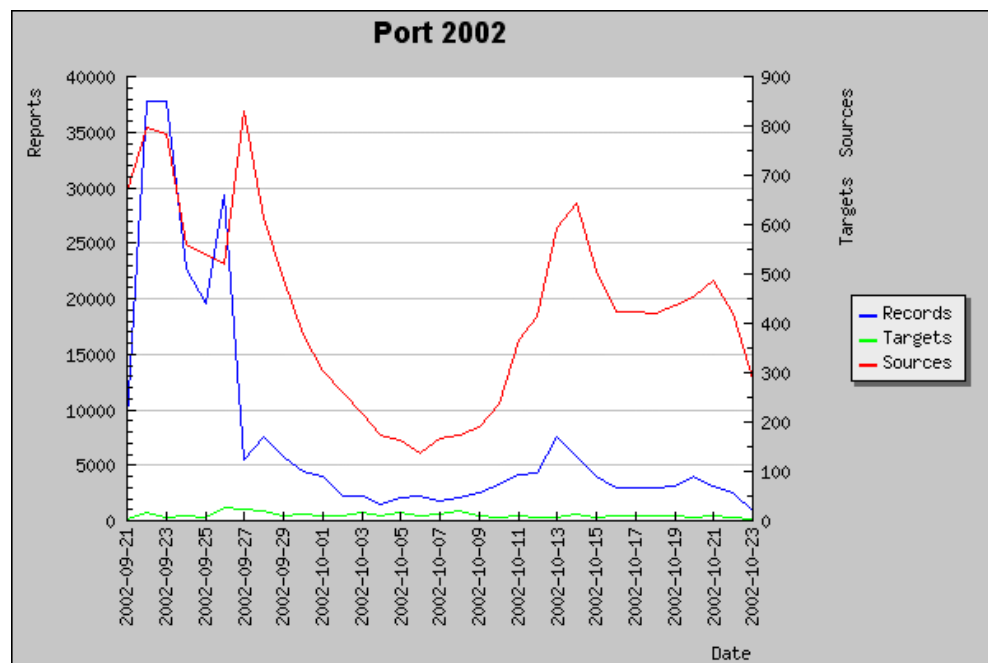
- **Detected a computer in the CS department talking on IPv6**
 - ◆ This is extremely rare traffic and represents a possible covert tunnel to the outside world
 - ◆ It turns out that the person doing this is on system staff and is in fact using this as a covert tunnel to his home computers

Detection of Worms on Real Network Data

◆ October 10, 2002

- **Detected several instances of slapper worm that were not identified by SNORT since they were variations of existing worm code**
- **Detected by *MINDS* anomaly detection algorithm since source and destination ports are the same but non-standard, and slow scan-like behavior for the source port**
- **Potentially detectable by SNORT using more general rules, but the false alarm rate will be too high**
- **Virus detection through anomalous behavior of infected machine**

Number of slapper worms on port 2002 reported in the World (Source www.incidents.org)





Detection of Worms on Real Network Data

◆ January 25 and January 26, 2003

◆ Even 24/48 hours after the “SQL Slammer/Sapphire” worm started, network connection related to the worm were ranked at the top of anomaly detection algorithm (24 hours after the “slammer” worm)

▪ The behavior of “slammer” worm is different from scanning activities, since infected machines target random hosts

score	srcIP	srcPort	dstIP	DstPort	protoc	flags	packets	bytes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
20826.69	128.171.137.62	1042	160.94.213.101	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
20344.83	128.171.137.62	1042	160.94.243.110	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
19295.82	128.171.137.62	1042	160.94.214.79	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18717.1	128.171.137.62	1042	160.94.155.47	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18147.16	128.171.137.62	1042	160.94.96.183	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
17484.13	128.171.137.62	1042	160.94.204.101	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
16715.61	128.171.137.62	1042	160.94.32.166	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
15973.26	128.171.137.62	1042	160.94.116.102	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
13084.25	128.171.137.62	1042	160.94.176.54	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
12797.73	128.171.137.62	1042	160.94.230.189	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
12428.45	128.171.137.62	1042	160.94.4.247	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
11245.21	128.171.137.62	1042	160.94.131.58	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9327.98	128.171.137.62	1042	160.94.148.135	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
7468.52	128.171.137.62	1042	160.94.182.91	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5489.69	128.171.137.62	1042	160.94.31.30	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5070.5	128.171.137.62	1042	160.94.180.233	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4558.72	128.171.137.62	1042	160.94.25.1	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4225.09	128.171.137.62	1042	160.94.133.143	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4170.72	128.171.137.62	1042	160.94.109.225	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2937.42	128.171.137.62	1042	160.94.135.75	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2458.61	128.171.137.62	1042	160.94.119.150	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1116.41	128.171.137.62	1042	160.94.187.255	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1035.17	128.171.137.62	1042	160.94.50.50	1434	17	16	[0,2)	[387,1264)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0



Detection of Worms on Real Network Data

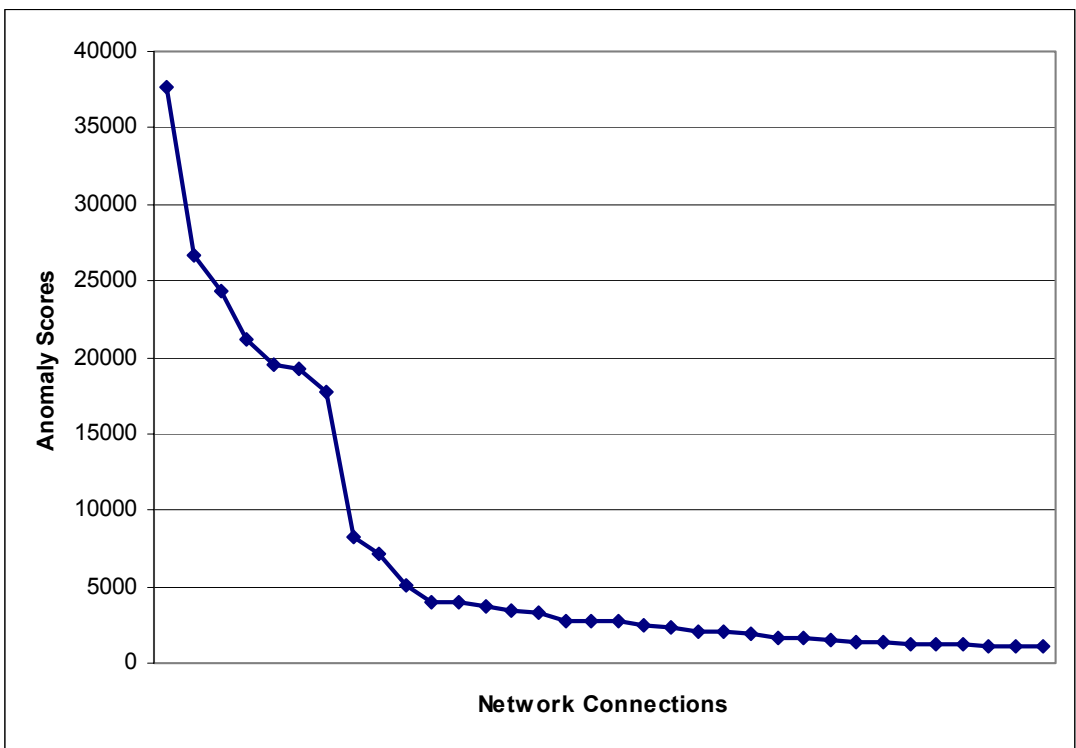
◆ January 26, 2003 (48 hours after the “slammer” worm)

score	srcIP	sPort	dstIP	dPort	protocc	flags	packets	bytes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
37674.69	63.150.X.253	1161	128.101.X.29	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.81	0	0.59	0	0	0	0	0
26676.62	63.150.X.253	1161	160.94.X.134	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.81	0	0.59	0	0	0	0	0
24323.55	63.150.X.253	1161	128.101.X.185	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.81	0	0.58	0	0	0	0	0
21169.49	63.150.X.253	1161	160.94.X.71	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.81	0	0.58	0	0	0	0	0
19525.31	63.150.X.253	1161	160.94.X.19	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.81	0	0.58	0	0	0	0	0
19235.39	63.150.X.253	1161	160.94.X.80	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.81	0	0.58	0	0	0	0	0
17679.1	63.150.X.253	1161	160.94.X.220	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.81	0	0.58	0	0	0	0	0
8183.58	63.150.X.253	1161	128.101.X.108	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.58	0	0	0	0	0
7142.98	63.150.X.253	1161	128.101.X.223	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
5139.01	63.150.X.253	1161	128.101.X.142	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
4048.49	142.150.Y.101	0	128.101.X.127	2048	1	16	[2,4)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
4008.35	200.250.Z.20	27016	128.101.X.116	4629	17	16	[2,4)	[0,1829)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
3657.23	202.175.Z.237	27016	128.101.X.116	4148	17	16	[2,4)	[0,1829)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
3450.9	63.150.X.253	1161	128.101.X.62	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
3327.98	63.150.X.253	1161	160.94.X.223	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
2796.13	63.150.X.253	1161	128.101.X.241	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
2693.88	142.150.Y.101	0	128.101.X.168	2048	1	16	[2,4)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
2683.05	63.150.X.253	1161	160.94.X.43	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
2444.16	142.150.Y.236	0	128.101.X.240	2048	1	16	[2,4)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
2385.42	142.150.Y.101	0	128.101.X.45	2048	1	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
2114.41	63.150.X.253	1161	160.94.X.183	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
2057.15	142.150.Y.101	0	128.101.X.161	2048	1	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
1919.54	142.150.Y.101	0	128.101.X.99	2048	1	16	[2,4)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
1634.38	142.150.Y.101	0	128.101.X.219	2048	1	16	[2,4)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
1596.26	63.150.X.253	1161	128.101.X.160	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
1513.96	142.150.Y.107	0	128.101.X.2	2048	1	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
1389.09	63.150.X.253	1161	128.101.X.30	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
1315.88	63.150.X.253	1161	128.101.X.40	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.82	0	0.57	0	0	0	0	0
1279.75	142.150.Y.103	0	128.101.X.202	2048	1	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
1237.97	63.150.X.253	1161	160.94.X.32	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
1180.82	63.150.X.253	1161	128.101.X.61	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0
1107.78	63.150.X.253	1161	160.94.X.154	1434	17	16	[0,2)	[0,1829)	0	0	0	0	0	0	0	0	0.83	0	0.56	0	0	0	0	0

Detection of Worms on Real Network Data

◆ January 26, 2003 (48 hours after the “slammer” worm)

Anomaly scores for “slammer” worm
after 48 hours of the real attack





Detection of Worms on Real Network Data

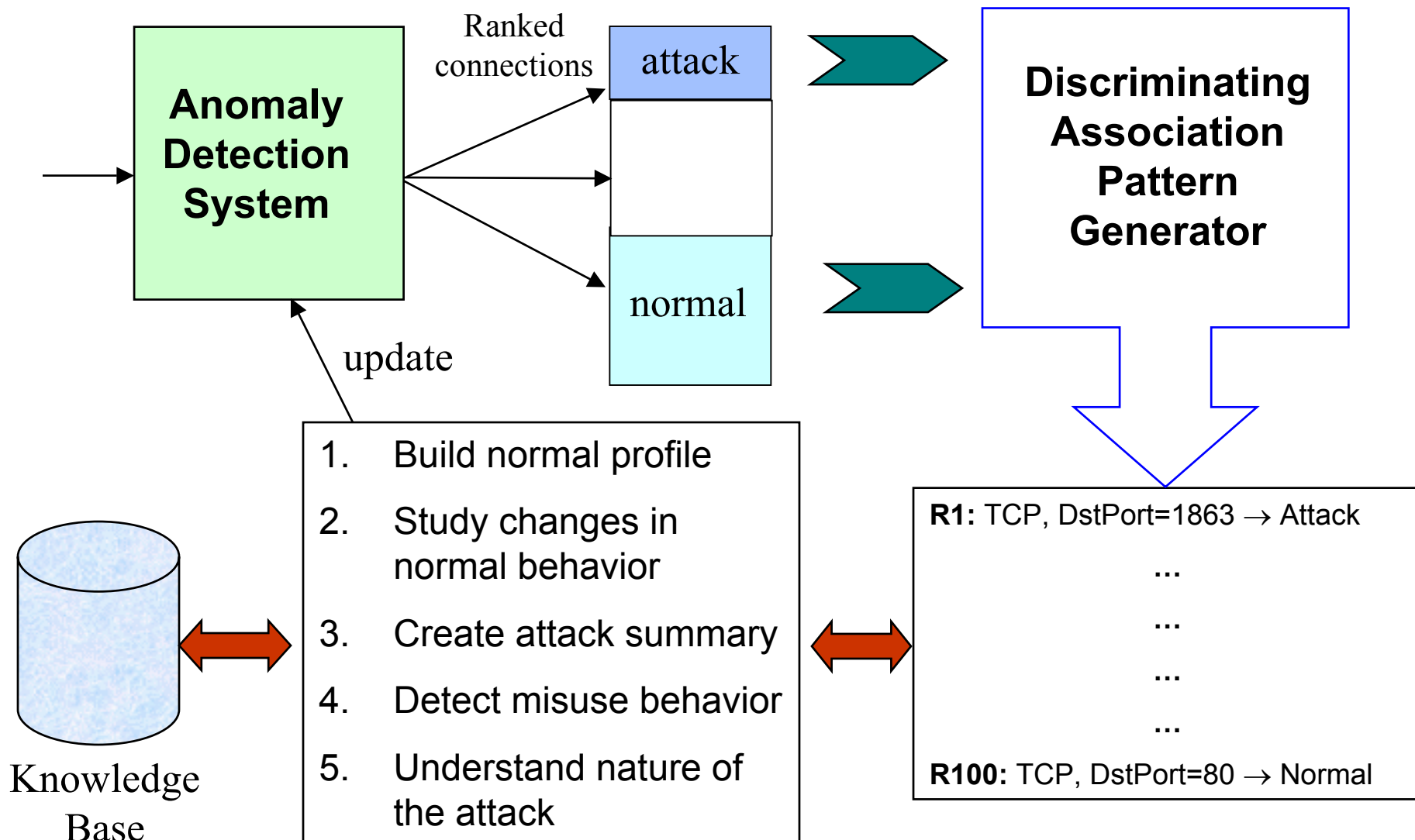
◆ February 6, 2003

- **Detected unsolicited ICMP ECHOREPLY messages to a computer previously infected with Stacheldract worm (a DDos agent)**
 - ◆ **Infected machine has been removed from the network**
 - ◆ **Other infected machines were still trying to talk to infected machine from our network**

- **Content-based attacks** (e.g. content of the packet)
 - ◆ **SNORT** is able to detect only those attacks with known signatures
 - ◆ Out of scope for **MINDS** anomaly/detection algorithms, since they do not use the content of the packets
- **Scanning activities**
 - ◆ **Fast scans**
 - **SNORT** and **MINDS** are equally good in identifying these attacks, provided that the scan satisfies the SNORT pre-specified thresholds (4 ports in 3 seconds)
 - ◆ **Slow scans**
 - **MINDS** anomaly/outlier detection identifies them better than **SNORT**, since **SNORT** has to increase time window which increases processing requirements

- **Policy violations** (e.g. rogue and unauthorized services)
 - ◆ **MINDS** anomaly/outlier detection algorithms are successful in detecting policy violations, since they are looking for unusual and suspicious network behavior
 - ◆ To detect these attacks **SNORT** has to have a rule for each specific unauthorized activity, which causes increase in the number of rules and therefore the memory requirements

MINDS - Framework for Mining Associations



Summarization of Anomalous Connections

- January 26, 2003 (48 hours after the Slammer worm)

score	srcIP	sPort	dstIP	dPort	protocc	flags	packets	bytes
37674.69	63.150.X.253	1161	128.101.X.29	1434	17	16	[0,2)	[0,1829]
26676.62	63.150.X.253	1161	160.94.X.134	1434	17	16	[0,2)	[0,1829]
24323.55	63.150.X.253	1161	128.101.X.185	1434	17	16	[0,2)	[0,1829]
21169.49	63.150.X.253	1161	160.94.X.71	1434	17	16	[0,2)	[0,1829]
19525.31	63.150.X.253	1161	160.94.X.19	1434	17	16	[0,2)	[0,1829]
19235.39	63.150.X.253	1161	160.94.X.80	1434	17	16	[0,2)	[0,1829]
17679.1	63.150.X.253	1161	160.94.X.220	1434	17	16	[0,2)	[0,1829]
8183.58	63.150.X.253	1161	128.101.X.108	1434	17	16	[0,2)	[0,1829]
7142.98	63.150.X.253	1161	128.101.X.223	1434	17	16	[0,2)	[0,1829]
5139.01	63.150.X.253	1161	128.101.X.142	1434	17	16	[0,2)	[0,1829]
4048.49	142.150.Y.101	0	128.101.X.127	2048	1	16	[2,4)	[0,1829]
4008.35	200.250.Z.20	27016	128.101.X.116	4629	17	16	[2,4)	[0,1829]
3657.23	202.175.Z.237	27016	128.101.X.116	4148	17	16	[2,4)	[0,1829]
3450.9	63.150.X.253	1161	128.101.X.62	1434	17	16	[0,2)	[0,1829]
3327.98	63.150.X.253	1161	160.94.X.223	1434	17	16	[0,2)	[0,1829]
2796.13	63.150.X.253	1161	128.101.X.241	1434	17	16	[0,2)	[0,1829]
2693.88	142.150.Y.101	0	128.101.X.168	2048	1	16	[2,4)	[0,1829]
2683.05	63.150.X.253	1161	160.94.X.43	1434	17	16	[0,2)	[0,1829]
2444.16	142.150.Y.236	0	128.101.X.240	2048	1	16	[2,4)	[0,1829]
2385.42	142.150.Y.101	0	128.101.X.45	2048	1	16	[0,2)	[0,1829]
2114.41	63.150.X.253	1161	160.94.X.183	1434	17	16	[0,2)	[0,1829]
2057.15	142.150.Y.101	0	128.101.X.161	2048	1	16	[0,2)	[0,1829]
1919.54	142.150.Y.101	0	128.101.X.99	2048	1	16	[2,4)	[0,1829]
1634.38	142.150.Y.101	0	128.101.X.219	2048	1	16	[2,4)	[0,1829]
1596.26	63.150.X.253	1161	128.101.X.160	1434	17	16	[0,2)	[0,1829]
1513.96	142.150.Y.107	0	128.101.X.2	2048	1	16	[0,2)	[0,1829]
1389.09	63.150.X.253	1161	128.101.X.30	1434	17	16	[0,2)	[0,1829]
1315.88	63.150.X.253	1161	128.101.X.40	1434	17	16	[0,2)	[0,1829]
1279.75	142.150.Y.103	0	128.101.X.202	2048	1	16	[0,2)	[0,1829]
1237.97	63.150.X.253	1161	160.94.X.32	1434	17	16	[0,2)	[0,1829]
1180.82	63.150.X.253	1161	128.101.X.61	1434	17	16	[0,2)	[0,1829]
1107.78	63.150.X.253	1161	160.94.X.154	1434	17	16	[0,2)	[0,1829]

Potential Rules:

1.

{Dest Port = 1434/UDP
#packets ∈ [0, 2)} -->
Highly anomalous behavior
(Slammer Worm)

2.

{Src IP = 142.150.Y.101,
Dest Port = 2048/ICMP
#bytes ∈ [0, 1829]} -->
Highly anomalous behavior
(ping – scan)

Discovered Real-life Association Patterns

Rule 1: SrcIP=IP1, DstPort=80, Protocol=TCP, Flag=SYN,
NoPackets: 3, NoBytes:120...180 (c1=256, c2 = 1)

Rule 2: SrcIP=IP1, DstIP=IP2, DstPort=80, Protocol=TCP,
Flag=SYN, NoPackets: 3, NoBytes: 120...180 (c1=177, c2 = 0)

- At first glance, Rule 1 appears to describe a Web scan
- Rule 2 indicates an attack on a specific machine
- Both rules together indicate that a scan is performed first, followed by an attack on a specific machine identified as vulnerable by the attacker

Discovered Real-life Association Patterns...(ctd)

DstIP=IP3, DstPort=8888, Protocol=TCP (c1=369, c2=0)

DstIP=IP3, DstPort=8888, Protocol=TCP, Flag=SYN (c1=291, c2=0)

- This pattern indicates an anomalously high number of TCP connections on port 8888 involving machine with IP address IP3
- Follow-up analysis of connections covered by the pattern indicates that this could be a machine running a variation of the Kazaa file-sharing protocol
- Having an unauthorized application increases the vulnerability of the system

Discovered Real-life Association Patterns...(ctd)

SrcIP=IP4, DstPort=27374, Protocol=TCP, Flag=SYN, NoPackets=4,
NoBytes=189...200 (c1=582, c2=2)

SrcIP=IP4, DstPort=12345, NoPackets=4, NoBytes=189...200 (c1=580,
c2=3)

SrcIP=IP5, DstPort=27374, Protocol=TCP, Flag=SYN, NoPackets=3,
NoBytes=144 (c1=694, c2=3)

.....

- This pattern indicates a large number of scans on ports 27374 (which is a signature for the SubSeven worm) and 12345 (which is a signature for NetBus worm)
- Further analysis showed that no fewer than five machines scanning for one or both of these ports in any time window

Discovered Real-life Association Patterns...(ctd)

`DstPort=6667, Protocol=TCP (c1=254, c2=1)`

- This pattern indicates an unusually large number of connections on port 6667 detected by the anomaly detector
- Port 6667 is where IRC (Internet Relay Chat) is typically run
- Further analysis reveals that there are many small packets from/to various IRC servers around the world
- Although IRC traffic is not unusual, the fact that it is flagged as anomalous is interesting
 - This might indicate that the IRC server has been taken down (by a DOS attack for example) or it is a rogue IRC server (it could be involved in some hacking activity)

Discovered Real-life Association Patterns...(ctd)

DstPort=1863, Protocol=TCP, Flag=0, NoPackets=1, NoBytes<139
(c1=498, c2=6)

DstPort=1863, Protocol=TCP, Flag=0 (c1=587, c2=6)

DstPort=1863, Protocol=TCP (c1=606, c2=8)

- This pattern indicates a large number of anomalous TCP connections on port 1863
- Further analysis reveals that the remote IP block is owned by Hotmail
- Flag=0 is unusual for TCP traffic


Conclusion



- Data mining based algorithms are capable of detecting intrusions that cannot be detected by state-of-the-art signature based methods
 - ◆ **SNORT** has static knowledge manually updated by human analysts
 - ◆ **MINDS** anomaly detection algorithms are adaptive in nature
 - ◆ **MINDS** anomaly detection algorithms can also be effective in detecting anomalous behavior originating from a compromised or infected machine

MINDS Research

- ◆ Defining normal behavior
- ◆ Feature extraction
- ◆ Similarity functions
- ◆ Outlier detection
- ◆ Result summarization
- ◆ Detection of attacks originating from multiple sites



Outsider attack

- ◆ Network intrusion



Insider attack

- ◆ Policy violation



**Worm/virus detection
after infection**

Future Work

- **Distributed Attacks coordinated from multiple locations**

