Computer Intrusion Detection

Lecture 4

Information Sources

Xiangyang Li

Outline

- General Requirements
- Host-based Information Sources
- Network-based Information Sources
- Two Sample Datasets
- Other Information Sources

General Considerations

What is the right information?

It should be able to reveal violation.

Where to collect?

Host versus network

Special instrument

Other Enduring Questions

- How much information is enough to allow you to accurately diagnose security problems without crippling the systems you are trying to protect?
- How do you manage the information collected to support any legal remedies you might want to pursue against attackers?
- How do you honor your responsibility to handle the information collected about users so that you stay within legal, regulatory, and ethical policy limits?

Format for Interoperability

- Common Intrusion Detection Framework (CIDF): Common Intrusion Specification Language (CISL)
- Intrusion Detection Message Exchange Format (IDMEF): XML-based
- Latest threat intelligence and sharing efforts: e.g., STIX/TAXII/CybOX

Vulnerable IDS

OVERLOAD MONITOR WITH EVENTS

SLOW PROCESSING

OVERLOAD DISK STORAGE

DOS ATTACKS AGAINST IDS

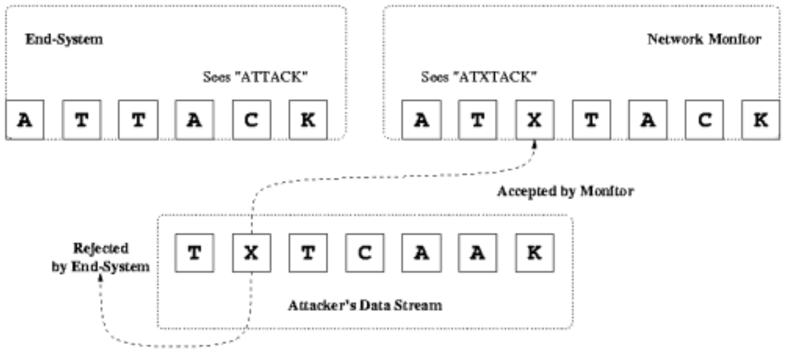


Figure 4: Insertion of the letter 'X'

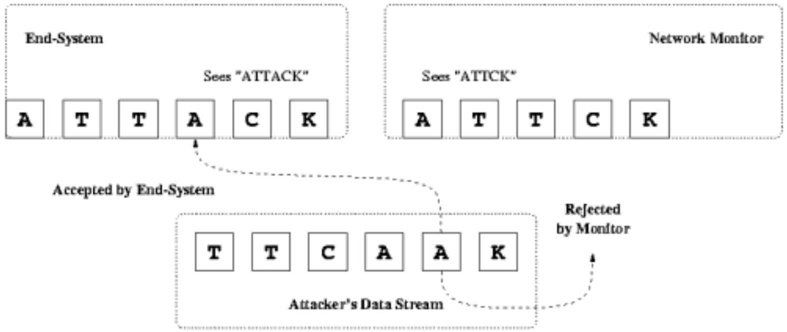


Figure 5: Evasion of the letter 'A'

The Insertion and Evasion Problems

Ptacek and Newsham (1998)

One Scenario

- If several internal routers exist between the Network-based IDS (NIDS) and destination host:
  - TTL may result in some packets reaching the NIDS but not the receiver.
  - Some packets are dropped by filtering routers after passing the NIDS.
- Other scenarios?

Generic Data Record

- A data record is a data point in space. We express a data point as a tuple:

  $\{X, A, Y\}$

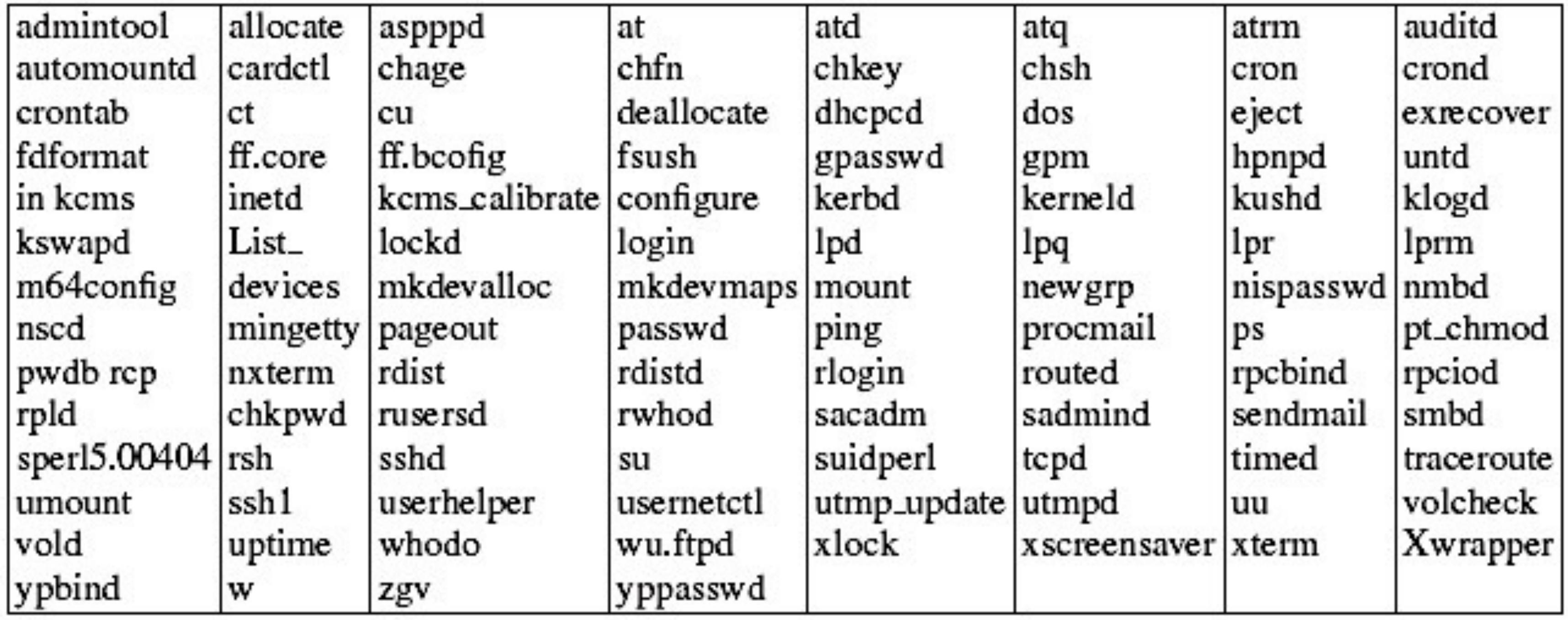
  where $X = (X_1, X_2, \ldots, X_p)$ represent the numeric attributes; $A = (A_1, A_2, \ldots, A_m)$ represents the nominal attributes; and Y is the nominal target variable representing the class of the data point.
- Each numeric predictor variable (attribute) X_i, $i \in \{1, \ldots, p\}$, can have a real value.
- Each nominal variable (attribute) A_i, $i \in \{1, \ldots, m\}$, can be a category value from the domain of this nominal attribute, DOM(A_i).

Outline

- General Requirements
- Host-based Information Sources
- Network-based Information Sources
- Two Sample Datasets
- Other Information Sources

Host-based Information Sources

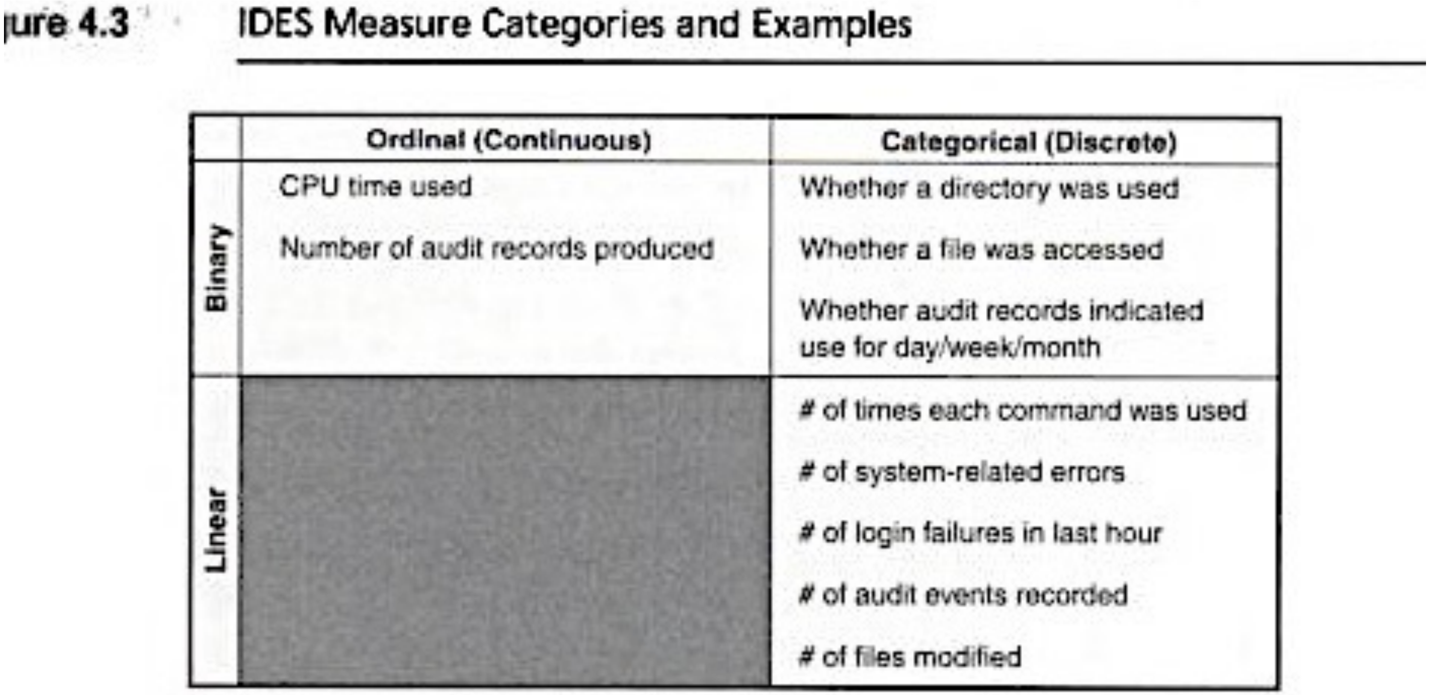
- Host-based Intrusion Detection Systems (HIDSs) analyze activities on a protected host by monitoring different sources of data that reside on that host.
  - OS audit trails
  - System logs
  - System calls
  - File access
  - Memory content
  - Application information



admintool	allocate	aspppd	at	atd	atq	atrm	auditd
automountd	cardctl	chage	chfn	chkey	chsh	cron	crond
crontab	ct	cu	deallocate	dhcpcd	dos	eject	exrecover
fdformat	ff.core	ff.bcofig	fsush	gpasswd	gpm	hpnpd	untd
in kcms	inetd	kcms_calibrate	configure	kerbd	kerneld	kushd	klogd
kswapd	List_	lockd	login	lpd	lpq	lpr	lprm
m64config	devices	mkdevalloc	mkdevmaps	mount	newgrp	nispasswd	nmbd
nscd	mingetty	pageout	passwd	ping	procmail	ps	pt_chmod
pwdb rcp	nxterm	rdist	rdistd	rlogin	routed	rpcbind	rpciod
rpld	chkpwd	rusersd	rwhod	sacadm	sadmind	sendmail	smbd
sperl5.00404	rsh	sshd	su	suidperl	tcpd	timed	traceroute
umount	ssh1	userhelper	usernetctl	utmp_update	utmpd	uu	volcheck
vold	uptime	whodo	wu.ftpd	xlock	xscreensaver	xterm	Xwrapper
ypbind	w	zgv	yppasswd				

Example Programs Monitored by HIDS

A.A. Ghorbani et al., *Network Intrusion Detection and Prevention: Concepts and Techniques*



ɹure 4.3 IDES Measure Categories and Examples

	Ordinal (Continuous)	Categorical (Discrete)
Binary	CPU time used Number of audit records produced	Whether a directory was used Whether a file was accessed Whether audit records indicated use for day/week/month
Linear		# of times each command was used # of system-related errors # of login failures in last hour # of audit events recorded # of files modified

Example: IDES Data

Operating System Audit Trails

- OS audit trails are generated by a specialized auditing subsystem included as part of the OS, to meet the requirements of the Trusted Computer System Evaluation Criteria TCSEC. (superseded by new DoD 8500.01E)
- A collection of information about system activities, at kernel (system call) and user (application) levels, are placed in chronological order into audit files.

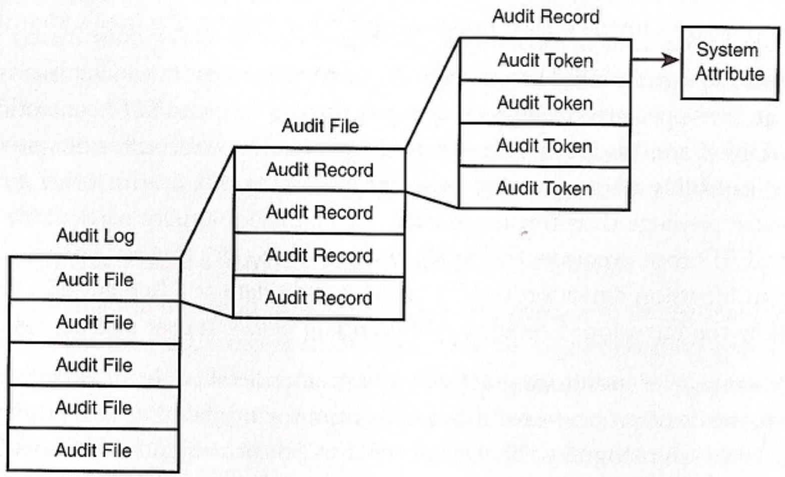
Operating System Audit Trails (cont.)

- Pros and cons
  - Protection
  - Finer-grained details
- Audit trail structuring
  - Self-contained
  - Distributed records for one event
- Problems in commercial audit systems
  - No documentation
  - No accurate documentation

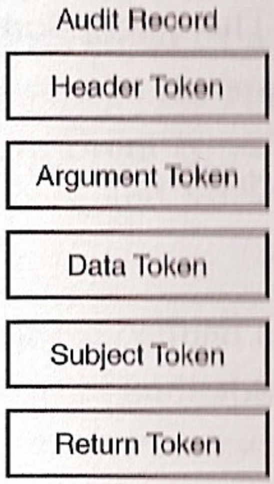
Sun Solaris BSM

- DoD "C2" level certification
- Audit files are binary.
- BSM has translation functions to translate audit events to kernel or user events.
- There are different event classes (>280).
- Audit trail management commands perform various functions.
  - Auditreduce to select events.
  - Praudit to translate the binary format to more readable ones.

Figure 3.1 Structure of Sun BSM Audit Data



Structure of Typical Sun BSM Audit Reco



BSM Audit Data Structure

BSM Audit Event Example

event	217
auid	-2
euid	0
egid	0
ruid	0
rgid	0
pid	96
sid	0
RemoteIP	0.0.0.0
time	897047263
error_message	91
process_error	0
retval	0

Windows NT

- Three types of system events are OS, security, and application events, in separate logs.
- The security log consists of security-relevant events, derived from TCSEC C2 definitions.
  - e.g. valid/invalid login/logoff, file use
  - Common Criteria and Mission Assurance Category (MAC) in the new DoD classification
- Each event record has a header, a description, and an optional additional data field.
- Administrator can manage the size of event log.

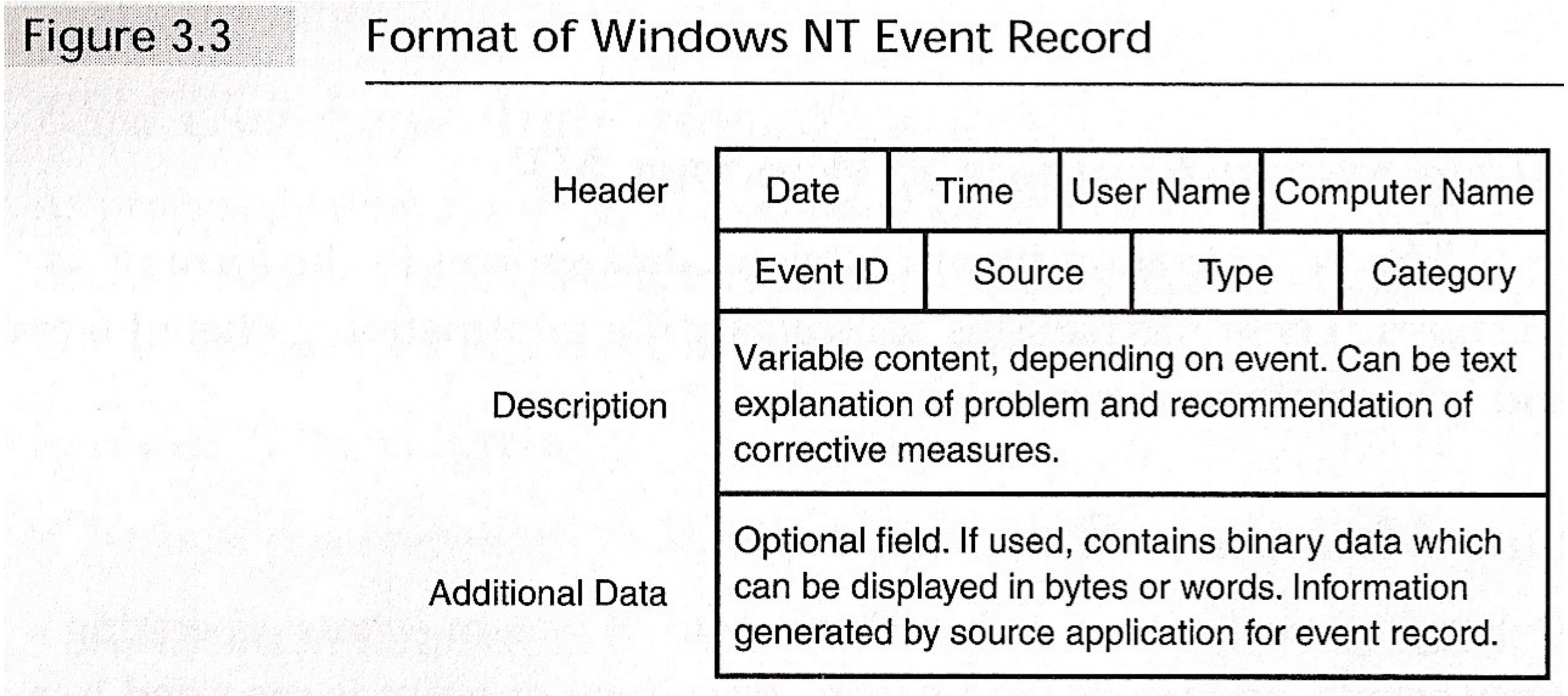


Figure 3.3 Format of Windows NT Event Record

Header	Date	Time	User Name	Computer Name
	Event ID	Source	Type	Category
Description	Variable content, depending on event. Can be text explanation of problem and recommendation of corrective measures.			
Additional Data	Optional field. If used, contains binary data which can be displayed in bytes or words. Information generated by source application for event record.			

Windows NT Audit Event Record

System Logs

THERE ARE OTHER LOG FILES FOR VARIOUS SYSTEMS EVENTS AND SETTINGS.

UNIX USES SYSLOG SERVICE WITH SYSLOGD DAEMON.

THESE LOGS ARE CONSIDERED LESS SECURE.

THESE ARE COMPLEMENT TO OS AUDIT.

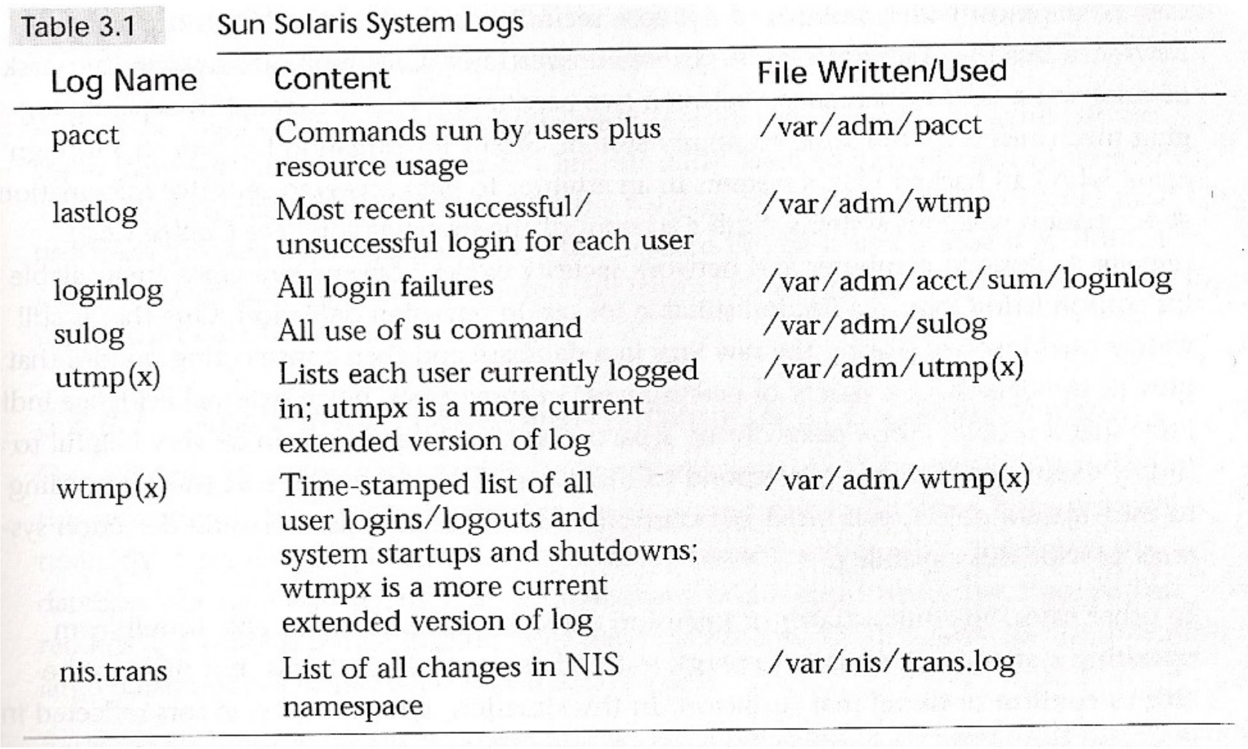


Table 3.1 Sun Solaris System Logs

Log Name	Content	File Written/Used
pacct	Commands run by users plus resource usage	/var/adm/pacct
lastlog	Most recent successful/ unsuccessful login for each user	/var/adm/wtmp
loginlog	All login failures	/var/adm/acct/sum/loginlog
sulog	All use of su command	/var/adm/sulog
utmp(x)	Lists each user currently logged in; utmpx is a more current extended version of log	/var/adm/utmp(x)
wtmp(x)	Time-stamped list of all user logins/logouts and system startups and shutdowns; wtmpx is a more current extended version of log	/var/adm/wtmp(x)
nis.trans	List of all changes in NIS namespace	/var/nis/trans.log

Sun Solaris Systems Logs

Application Information

- In modern systems, application logs may often represent the only available user-level abstraction of system activity.
- Development of object-oriented and distributed systems enhances this.
- OS audit mechanisms support the generation of application-level audit entries, but few include application with auditing features.

Database Systems

- Volume may be more an issue.
  - Compression and archival
  - Audit reduction
  - Granularity of audit control, e.g. switch on one event type vs one groups of events
- Temporal discrepancy can be induced in time due to the application-level auditing.
- Similar trade-off issue exists with level of abstraction.
  - Can we just use the transaction log?
  - Composition and fusion

WWW Servers

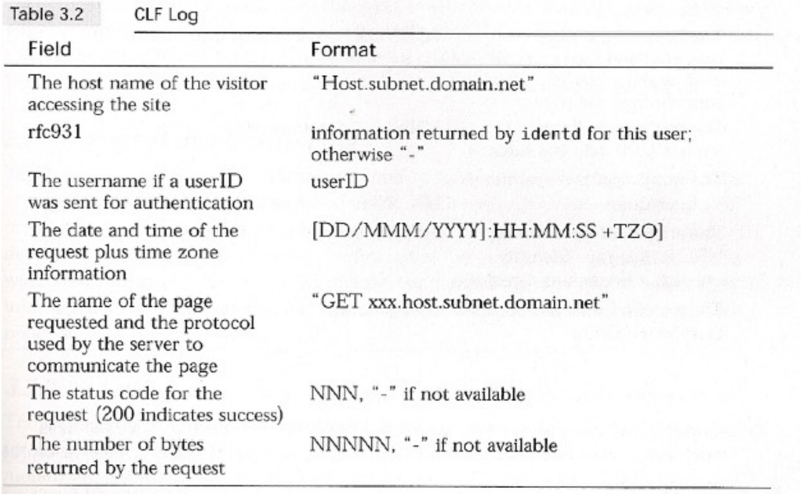


Table 3.2 CLF Log

Field	Format
The host name of the visitor accessing the site	"Host.subnet.domain.net"
rfc931	information returned by `identd` for this user; otherwise "-"
The username if a userID was sent for authentication	userID
The date and time of the request plus time zone information	[DD/MMM/YYYY]:HH:MM:SS +TZO]
The name of the page requested and the protocol used by the server to communicate the page	"GET xxx.host.subnet.domain.net"
The status code for the request (200 indicates success)	NNN, "-" if not available
The number of bytes returned by the request	NNNNN, "-" if not available

- Two types of access log formats are Common Log Format (CLF) and an extension on the specific web server.

Target-based Monitoring

- Monitoring mechanism is designed to collect information about the most critical or valuable (or of interest) objects in the system.
  - cryptographic integrity checkers, e.g., Tripwire
- In UNIX all items of interests to users can be represented as files, in structures called inodes.
  - network connection
  - device
  - process

Outline

- General Requirements
- Host-based Information Sources
- Network-based Information Sources
- Two Sample Datasets
- Other Information Sources

Network-based Information Sources

- Low to no performance cost to monitor
- Monitor is transparent to the users so safe to certain extent.
- Network traffic is necessary to identify certain attacks such as malformed packets and DDoS.

Different Features of Network Traffic

Packet header values

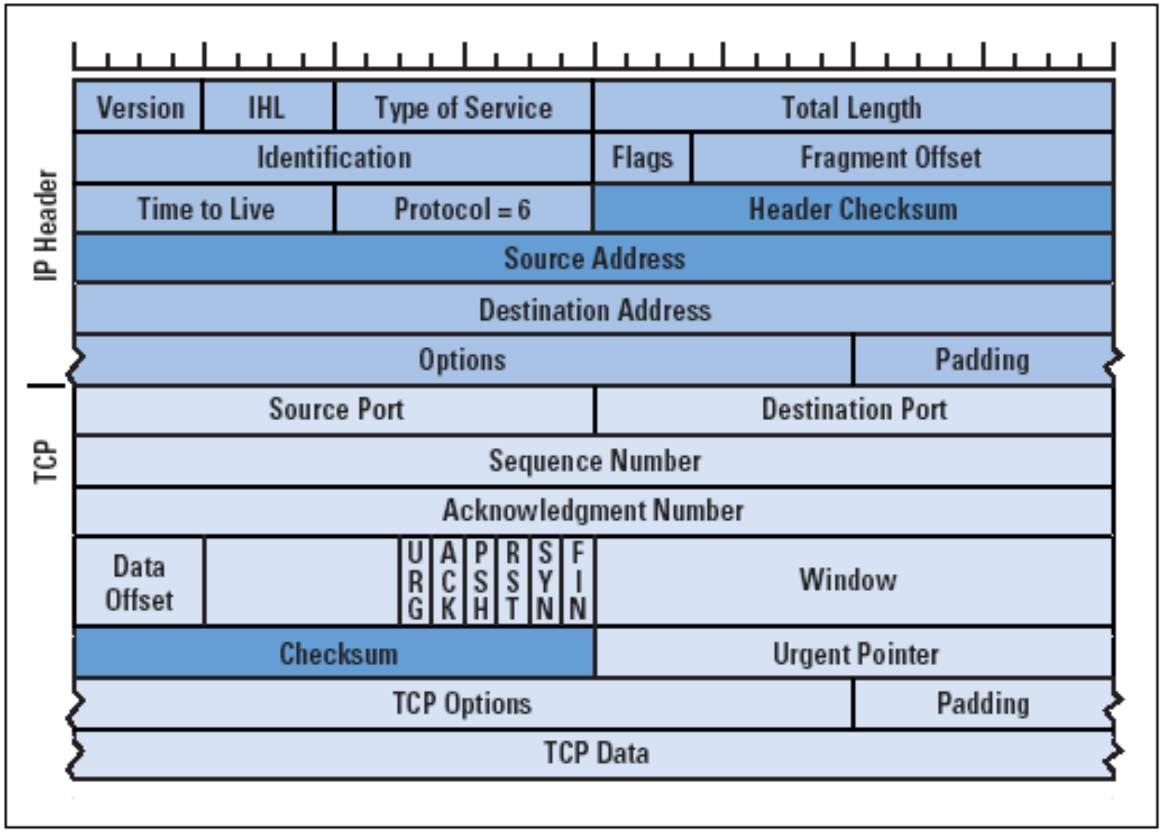
Mahoney (2002)

TCP Sessions

Lee (2002)

Behavioral features

Early (2003)



IP Header

Version | IHL | Type of Service | Total Length

Identification | Flags | Fragment Offset

Time to Live | Protocol = 6 | Header Checksum

Source Address

Destination Address

Options | Padding

TCP

Source Port | Destination Port

Sequence Number

Acknowledgment Number

Data Offset | URG | ACK | PSH | RST | SYN | FIN | Window

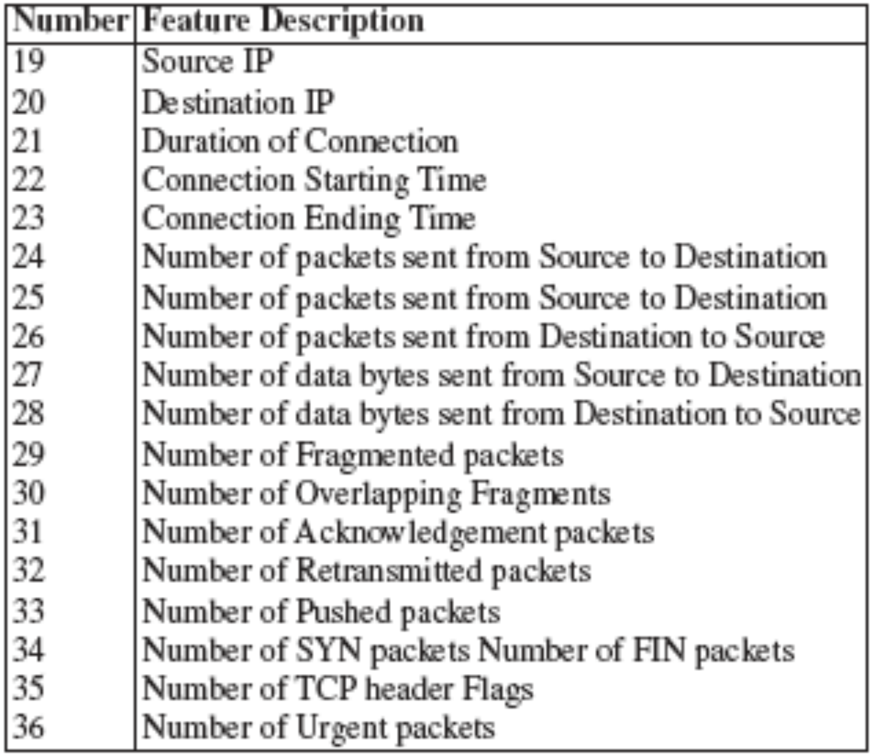
Checksum | Urgent Pointer

TCP Options | Padding

TCP Data

(https://erlerobotics.gitbooks.io/erle-robotics-introduction-to-linux-networking/content/introduction_to_network/tcp_and_packets.html)

Packet Headers



Number	Feature Description
19	Source IP
20	Destination IP
21	Duration of Connection
22	Connection Starting Time
23	Connection Ending Time
24	Number of packets sent from Source to Destination
25	Number of packets sent from Source to Destination
26	Number of packets sent from Destination to Source
27	Number of data bytes sent from Source to Destination
28	Number of data bytes sent from Destination to Source
29	Number of Fragmented packets
30	Number of Overlapping Fragments
31	Number of Acknowledgement packets
32	Number of Retransmitted packets
33	Number of Pushed packets
34	Number of SYN packets Number of FIN packets
35	Number of TCP header Flags
36	Number of Urgent packets

Network Connection Data

Packet Capture

Windows packet capture options

Microsoft Network Monitor

WinPcap

WinDump

UNIX packet capture options

Libpcap

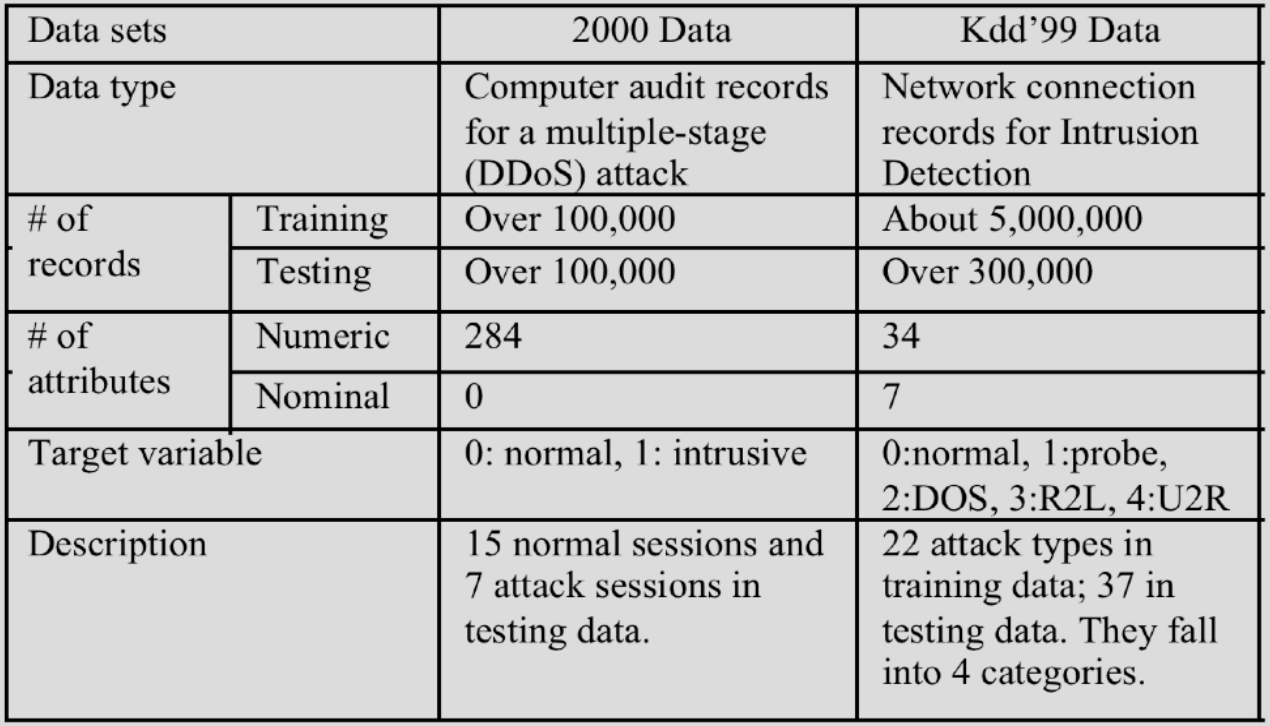
Tcpdump

Other packet filters

Outline

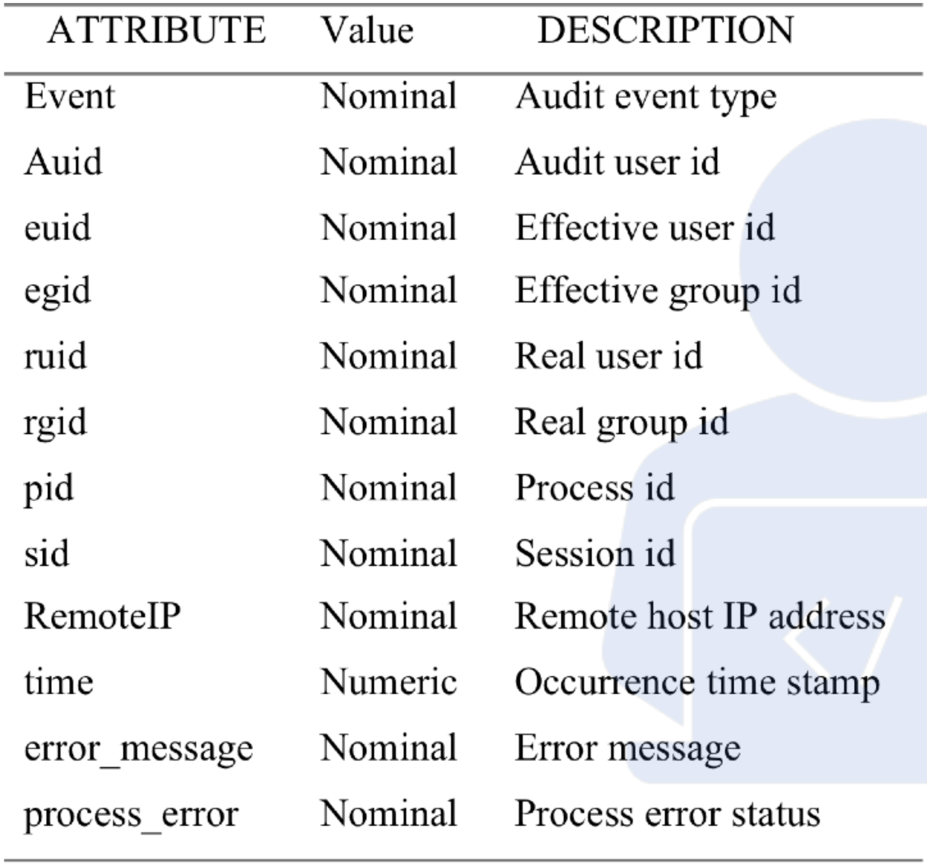
- General Requirements
- Host-based Information Sources
- Network-based Information Sources
- Two Sample Datasets
- Other Information Sources

Two Popular Data Sets



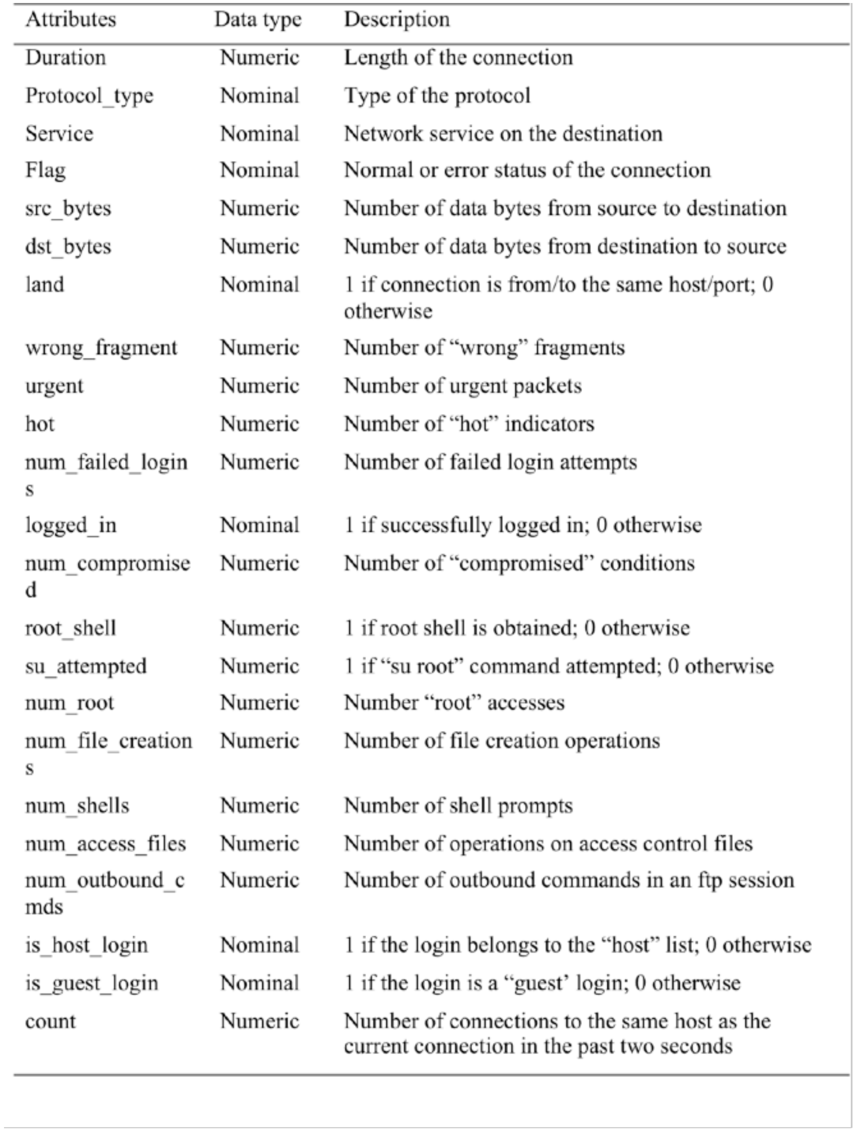
Data sets		2000 Data	Kdd’99 Data
Data type		Computer audit records for a multiple-stage (DDoS) attack	Network connection records for Intrusion Detection
# of records	Training	Over 100,000	About 5,000,000
	Testing	Over 100,000	Over 300,000
# of attributes	Numeric	284	34
	Nominal	0	7
Target variable		0: normal, 1: intrusive	0:normal, 1:probe, 2:DOS, 3:R2L, 4:U2R
Description		15 normal sessions and 7 attack sessions in testing data.	22 attack types in training data; 37 in testing data. They fall into 4 categories.

Audit Data Record in 2000 Data



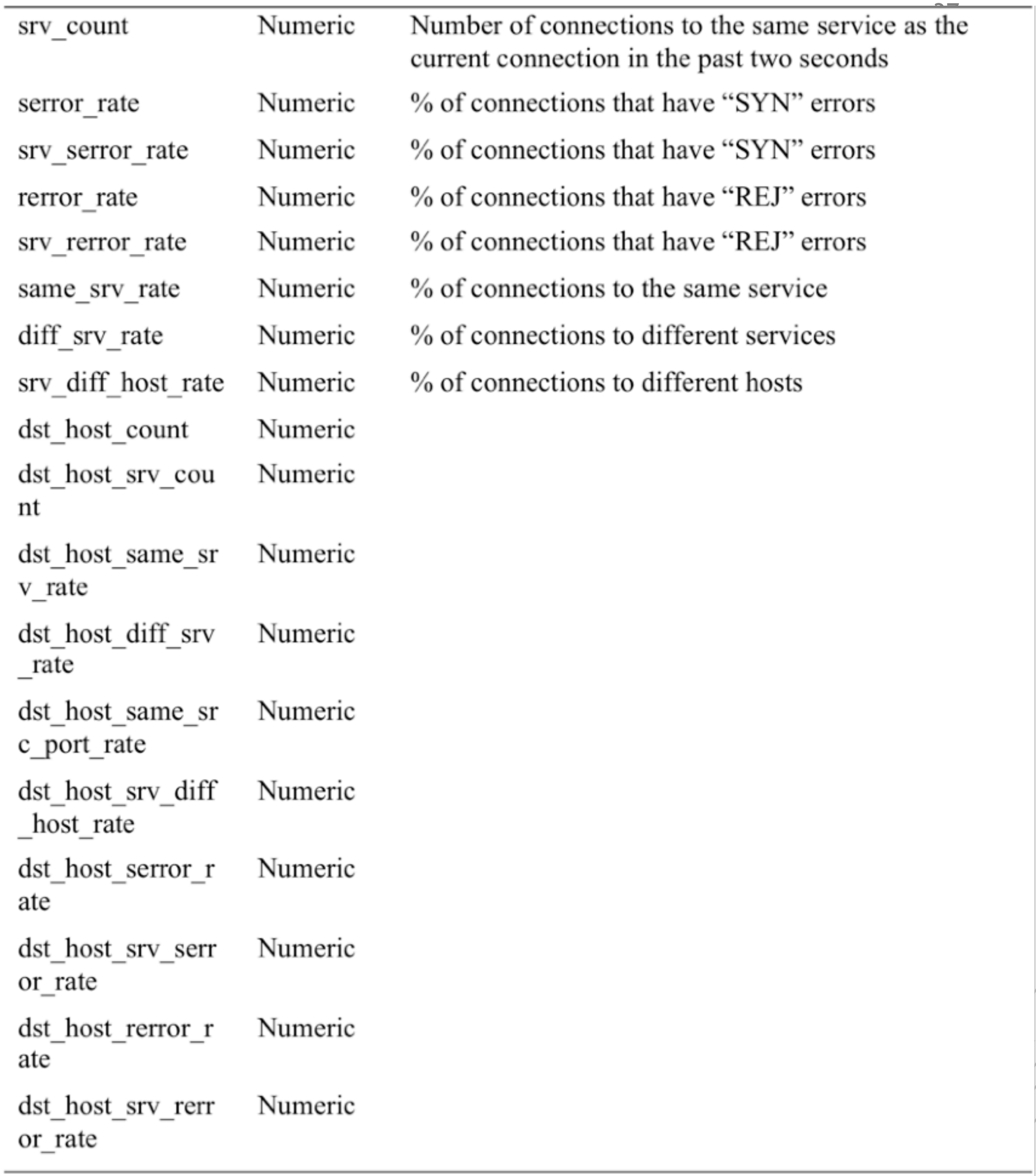
ATTRIBUTE	Value	DESCRIPTION
Event	Nominal	Audit event type
Auid	Nominal	Audit user id
euid	Nominal	Effective user id
egid	Nominal	Effective group id
ruid	Nominal	Real user id
rgid	Nominal	Real group id
pid	Nominal	Process id
sid	Nominal	Session id
RemoteIP	Nominal	Remote host IP address
time	Numeric	Occurrence time stamp
error_message	Nominal	Error message
process_error	Nominal	Process error status

Network Connect Record in Kdd99



Attributes	Data type	Description
Duration	Numeric	Length of the connection
Protocol_type	Nominal	Type of the protocol
Service	Nominal	Network service on the destination
Flag	Nominal	Normal or error status of the connection
src_bytes	Numeric	Number of data bytes from source to destination
dst_bytes	Numeric	Number of data bytes from destination to source
land	Nominal	1 if connection is from/to the same host/port; 0 otherwise
wrong_fragment	Numeric	Number of “wrong” fragments
urgent	Numeric	Number of urgent packets
hot	Numeric	Number of “hot” indicators
num_failed_logins	Numeric	Number of failed login attempts
logged_in	Nominal	1 if successfully logged in; 0 otherwise
num_compromised	Numeric	Number of “compromised” conditions
root_shell	Numeric	1 if root shell is obtained; 0 otherwise
su_attempted	Numeric	1 if “su root” command attempted; 0 otherwise
num_root	Numeric	Number “root” accesses
num_file_creations	Numeric	Number of file creation operations
num_shells	Numeric	Number of shell prompts
num_access_files	Numeric	Number of operations on access control files
num_outbound_cmds	Numeric	Number of outbound commands in an ftp session
is_host_login	Nominal	1 if the login belongs to the “host” list; 0 otherwise
is_guest_login	Nominal	1 if the login is a “guest’ login; 0 otherwise
count	Numeric	Number of connections to the same host as the current connection in the past two seconds

Network Connect Record in Kdd99 (cont.)

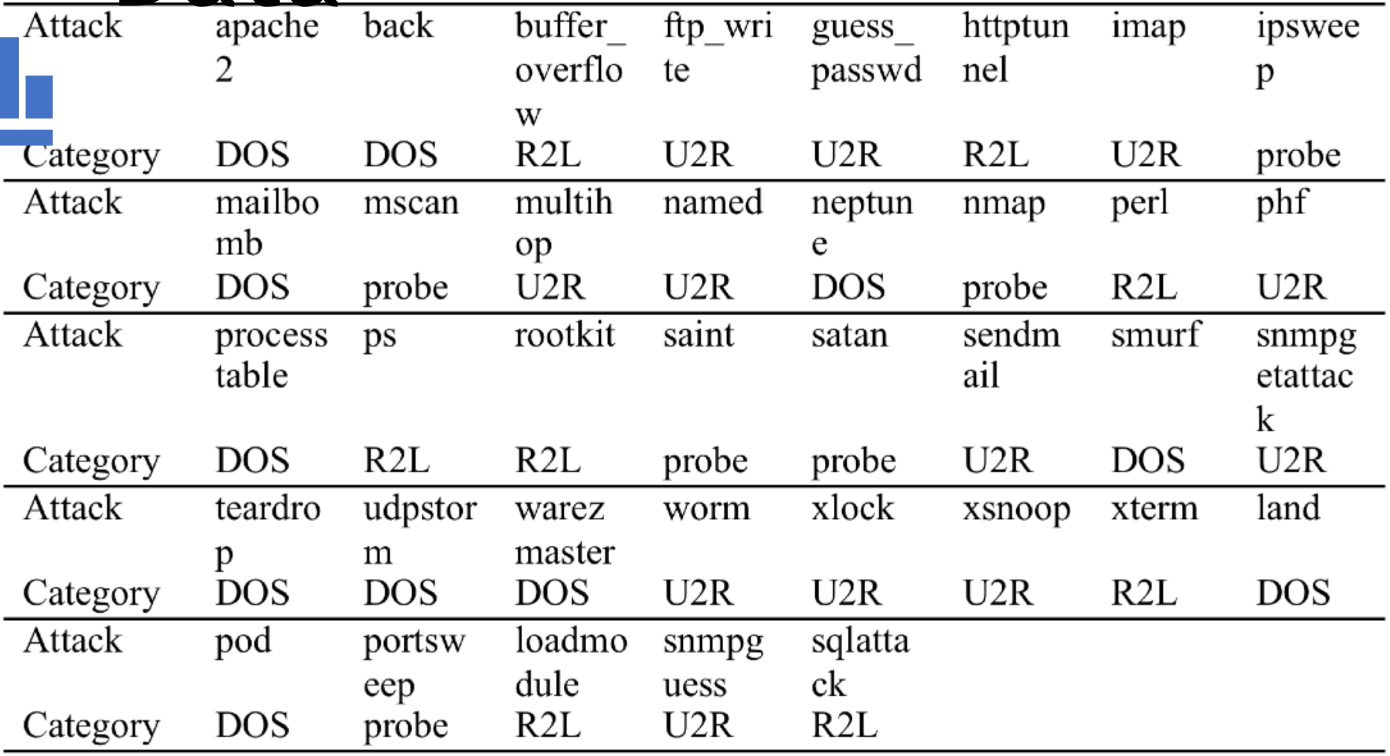


srv_count	Numeric	Number of connections to the same service as the current connection in the past two seconds
serror_rate	Numeric	% of connections that have “SYN” errors
srv_serror_rate	Numeric	% of connections that have “SYN” errors
rerror_rate	Numeric	% of connections that have “REJ” errors
srv_rerror_rate	Numeric	% of connections that have “REJ” errors
same_srv_rate	Numeric	% of connections to the same service
diff_srv_rate	Numeric	% of connections to different services
srv_diff_host_rate	Numeric	% of connections to different hosts
dst_host_count	Numeric	
dst_host_srv_count	Numeric	
dst_host_same_srv_rate	Numeric	
dst_host_diff_srv_rate	Numeric	
dst_host_same_src_port_rate	Numeric	
dst_host_srv_diff_host_rate	Numeric	
dst_host_serror_rate	Numeric	
dst_host_srv_serror_rate	Numeric	
dst_host_rerror_rate	Numeric	
dst_host_srv_rerror_rate	Numeric	

Attack Categories In KDD’99 Training Data

Attack	back	buffer_overflow	ftp_write	guess_passwd	imap	ipsweep	land	loadmodule
Category	DOS	U2R	R2L	R2L	R2L	probe	DOS	U2R
Attack	perl	phf	pod	portsweep	rootkit	satan	smurf	spy
Category	U2R	R2L	DOS	probe	U2R	probe	DOS	R2L
Attack	neptune	nmap	warezclient	warezmaster	multihop	teardrop		
Category	DOS	probe	R2L	R2L	R2L	DOS		

Attack Categories In KDD’99 **Testing** Data



Attack	apache2	back	buffer_overflow	ftp_write	guess_passwd	httptunnel	imap	ipsweep
Category	DOS	DOS	R2L	U2R	U2R	R2L	U2R	probe
Attack	mailbomb	mscan	multihop	named	neptune	nmap	perl	phf
Category	DOS	probe	U2R	U2R	DOS	probe	R2L	U2R
Attack	processtable	ps	rootkit	saint	satan	sendmail	smurf	snmpgetattack
Category	DOS	R2L	R2L	probe	probe	U2R	DOS	U2R
Attack	teardrop	udpstorm	warezmaster	worm	xlock	xsnoop	xterm	land
Category	DOS	DOS	DOS	U2R	U2R	U2R	R2L	DOS
Attack	pod	portsweep	loadmodule	snmpguess	sqlattack			
Category	DOS	probe	R2L	U2R	R2L			

Outline

General Requirements

Host-based Information Sources

Network-based Information Sources

Two Sample Datasets

Other Information Sources

Information from SNMP and Others

- Simple Network Management Protocol (SNMP) is an "Internet-standard protocol for managing devices on IP networks. Devices that typically support SNMP include routers, switches, servers, workstations, printers, modem racks, and more."
- In SNMP, the so-called Network Management Systems (NMSs) monitor the network status passively and provide information about the network traffic statistics.
- Many other network devices also provide relevant information.

Information from Other Security Products

- Many firewalls, information assurance systems, access control systems, and other security devices generate activities.
- Integrating and analyzing event logs from other components of the system security infrastructure plays an important role.

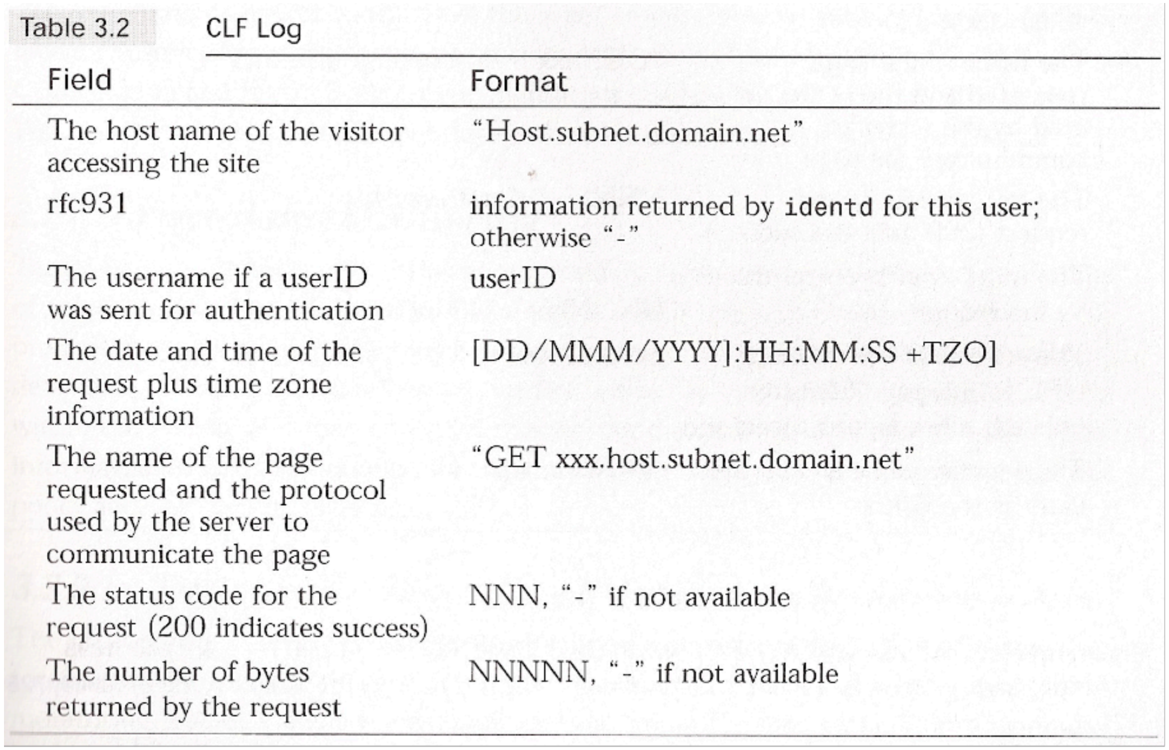


Table 3.2 CLF Log

Field	Format
The host name of the visitor accessing the site	“Host.subnet.domain.net”
rfc931	information returned by `identd` for this user; otherwise “-”
The username if a userID was sent for authentication	userID
The date and time of the request plus time zone information	[DD/MMM/YYYY]:HH:MM:SS +TZO]
The name of the page requested and the protocol used by the server to communicate the page	“GET xxx.host.subnet.domain.net”
The status code for the request (200 indicates success)	NNN, “-” if not available
The number of bytes returned by the request	NNNNN, “-” if not available

Firewall-1 Log Files

Honeypots and Honeynets

- Real or virtual system attractive to hackers, without other practical use.
- These systems entice attackers to break in.
- Observe and record all activities.
- Then use the collected information to develop stronger defense and detection.

Other System Components

- Intrusion detection can be aided with physical access control to identify masquerader.
  - Whether the user in question is on the premises?
- Out-of-band information source
  - Not input from computer/network, but from human or other systems such as telephone records