

The Quantitative Comparison of Computer Networks

By

S TERRY BRUGGER

B.S. (Purdue University) 1997

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Matt Bishop

Shyhtsun Wu

Prasant Mohapatra

Committee in Charge

2009

© S Terry Brugger, 2009. All rights reserved.

S Terry Brugger
June 2009
Computer Science

The Quantitative Comparison of Computer Networks

Abstract

Measurement is the basis of the scientific method. If we can measure two things, we can compare them. Until we can do this with computer networks, we will not have a field of network science.

Being able to compare two traces from computer networks in a quantifiable, holistic, and meaningful way provides us better insight into the network, and enables a range of applications for forensics, administration, application and protocol development, traffic generation, and general network intelligence.

In this work we propose a methodology for such a quantifiable comparison of two network traces, be they from different networks, or the same network at different times. This methodology is intended to be the primary contribution in this work to advance the state of the art. The basic approach of the methodology is to define all the possible relevant metrics available in a trace, extract those metrics from both traces, find the similarity for each individual metric between the two traces, and combine the similarities into an overall similarity metric by summing together their weighted values. The proper way to determine these weights was inconclusive.

The bulk of this work – volume wise – is dedicated to the construction and application of a specific instantiation of this methodology on IPv4 traffic. We begin by testing against artificial traffic, then show its applicability using a large number of publicly available network traces.

The key contributions of this work are a set of characteristics for comparing IPv4 traffic, and methods for modeling and comparing network characteristics.

To the Kara Brugger Home for Wayward Computer Scientists

TABLE OF CONTENTS

Abstract	ii
Dedication	iii
Table of Contents	iv
List of Tables	ix
List of Figures	xvi
Preface	xxx
Acknowledgments	xxxi
1 Introduction	1
1. Problem statement	1
2. Approach	3
3. Outline	3
4. Conclusion	4
2 The Need for Network Comparison	5
1. Usefulness of a More Holistic Comparison	5
1. Network Forensics	5
2. Network Administration	6
3. Network Application Development	6
4. Network Protocol Development	6
5. Network Traffic Generation	7
6. Network Intelligence	7
2. Previous work	8
3 Proposed Methodology	15
1. Methodology summary	15
1. Derivation of characteristics and weights	15
2. Comparison of network traces	17
2. Detailed Methodology	17
1. Derivation of Characteristics and Weights	17
2. Comparison of IP Network Traces	30
3. Success metrics	33
4 Initial steps	34
1. Characteristics for comparing IPv4 networks	34
2. Base cases	35
3. Initial test data	37
5 Single Value and Discrete Metrics on Initial Base Cases	38
1. The trouble with count and discrete metrics	38
2. Single value and discrete metrics	40
1. Count metrics	41
2. Ratio metrics	42
3. Discrete metrics	44
3. Normalized similarity	47
4. Scaled similarity	47

6	Connection Metrics	51
1.	Introduction	51
2.	Viewpoint	52
3.	Ambiguity in defining TCP connections	54
4.	Ambiguous TCP connection metrics	55
1.	Number of data or control packets	56
2.	Resend rate	56
3.	Number of establishment errors	57
5.	Sorting by time	58
6.	Logical Connections	58
7	Initial Continuous Characteristic Tests	60
1.	Measuring	61
2.	Modeling continuous distributions	61
1.	Definitions	65
2.	Modeling ordered continuous characteristics	66
3.	Modeling sorted continuous characteristics	66
4.	Modeling non-keyed sorted continuous characteristics	67
3.	Normalized Similarity	67
4.	Scaled Similarity	68
8	Tests on real data	69
1.	Usable network traces	69
1.	Linode	69
2.	DSL1	70
3.	Dartmouth campus	71
4.	LBNL	72
5.	SOTM27	72
6.	SIGCOMM2004	73
7.	AMES Internet Exchange OC-48	74
8.	DefCon 10	75
2.	Real basecases	75
3.	Normalized similarity	79
4.	Weighting factors and Scaled similarity	80
1.	Training cases	80
2.	Validation	81
3.	Test cases	86
9	Analysis	89
1.	Useful characteristics	89
1.	Unpriv connections connection time rate	90
2.	Packet Destination IP	90
3.	FINs connection time rate	90
4.	InterPacket delta	92
2.	Non-useful characteristics	92
3.	Reports	95
4.	Evaluation	95

10	Conclusion	97
1.	Summary of Findings	97
1.	Characteristics	97
2.	Extracting characteristics	98
3.	Modeling characteristics	98
4.	Comparing characteristics	99
2.	Future Work	100
	Bibliography	102

Appendices

A	Similarity Calculation Code	108
B	Characteristics for comparing IPv4 networks	113
1.	Single value metrics	113
2.	Discrete metrics	114
3.	Continuous metrics	114
1.	Per packet	114
2.	Per packet, time rates	115
3.	Per packet, packet rates	116
4.	Per connection initiation	116
5.	Per connection initiation, connection rates	117
6.	Per connection close	117
7.	Per IP	119
C	Initial test data	120
1.	Base case 1	120
2.	Base case 2	120
3.	Base case 3	121
4.	Base case 4	122
5.	Base case 5	123
D	Count and Discrete Metrics on Initial Base Cases	126
1.	Count characteristics	126
1.	Measuring	127
2.	Normalized Similarity	129
3.	Scaled Similarity	130
2.	Discrete characteristics	133
1.	Measuring	134
2.	Dealing with IP address differences	135
3.	Non-linear versus linear weighting	136
E	More problems with linear scaling	143
F	Modeling Interarrival time of packets	147
G	Modeling TTL	158

H	Details of Initial Continuous Characteristic Tests	169
1.	Modeling continuous distributions	169
1.1.	Per packet	169
1.2.	Per packet, time rates	179
1.3.	Per packet, packet rates	195
1.4.	Per connection initiation	197
1.5.	Per connection initiation, connection rates	204
1.6.	Per connection close	215
1.7.	Per IP	239
2.	Normalized Similarity	243
3.	Scaled Similarity	250
I	Notes on real data	255
1.	Internet Traffic Archive	255
2.	UNC/FORTH	255
3.	MAWI	255
4.	Crawdad	256
4.1.	Dataset: dartmouth/campus	256
4.2.	Dataset: microsoft/osdi2006	257
4.3.	Dataset: pdx/vwave	258
4.4.	Dataset: ucsb/ietf2005/wireless	258
4.5.	Dataset: ucsd/sigcomm2001	259
4.6.	Dataset: uw/sigcomm2004	260
5.	UMass trace repository	260
5.1.	Dataset: uprm-wireless	261
5.2.	Dataset: web ident	261
6.	UNC Mobile	262
7.	CAIDA (DatCat)	262
7.1.	Dataset: AMES Internet Exchange (OC-48)	262
7.2.	Dataset: passive-2007/ampath-oc12	263
7.3.	Dataset: SOTM27	264
7.4.	Dataset: Numerous DNS root server traces	264
7.5.	Dataset: KAIST-KOREN 1 Gigabit Ethernet Trace	264
7.6.	Dataset: LBNL	265
8.	DefCon CTF	274
8.1.	Shmoo data	274
8.2.	Dataset: DefCon 9	274
9.	DSL1	275
10.	Linode	275
11.	DSL2	275
12.	Data timeline	276
13.	Data Constraints	276
13.1.	LBNL	276
13.2.	DefCon10	277
13.3.	Ames	277
13.4.	Dartmouth	278
13.5.	DSL1	280
13.6.	Linode	281
13.7.	SOTM27	282
13.8.	SIGCOMM2004	283
14.	Unused descriptions	284
14.1.	MAWI	284
14.2.	Ampath OC-12	285

J	Real Data Details	287
K	Real data distributions	296
L	Real data normalized similarities	418
M	Analysis Details	648
1.	Reports	650
1.	Report for basecase 1, testing pair 2	650
2.	Report for basecase 3, testing pair 2	654
3.	Report for basecase 3, testing pair 3	658
4.	Report for basecase 8, testing pair 3	662

LIST OF TABLES

Table 3.1	Distributions for characteristics of the first trace for each base pair of traces	20
Table 3.2	Unscaled distributions for characteristics of the second trace for each base pair of traces	21
Table 3.3	Scaling factors for each pair of traces	21
Table 3.4	Scaled distributions for characteristics of the second trace for each base pair of traces	22
Table 3.5	Per characteristic similarity values for each of the pairs of traces.	25
Table 3.6	Per characteristic weighting factors (wf) for each of the pairs of traces to bring the corresponding similarity to 0.9 for pair 1 and 0.25 for pair 2, followed by the average weighting factor, and then the scaled similarities for each characteristic in each pair of traces.	25
Table 3.7	Distributions on the first traces of each pair for the second-order characteristics	28
Table 3.8	Unscaled distributions on the second traces of each pair for the second-order characteristics	29
Table 3.9	Scaled distributions on the second traces of each pair for the second-order characteristics	29
Table 3.10	Normalized similarities of second-order characteristics	30
Table 3.11	Per characteristic weighting factors (wf) for each of the pairs of traces to bring the corresponding similarity to 0.9 for pair 1 and 0.25 for pair 2, followed by the average weighting factor, and then the scaled similarities for each second-order characteristic in each pair of traces.	30
Table 3.12	Distributions for all characteristics from all traces	32
Table 3.13	Scaled similarities of all characteristics	33
Table 5.1	The normalized similarities of count and discrete characteristics for basecases 1 through 5.	48
Table 5.2	The weighting factors for the singular and discrete characteristics found with linear regression, using the goal similarities of {1.0, 1.0, 0.9, 0.8, 0.1}.	49
Table 5.3	The goal and calculated similarities for the first five basecases using the singular and discrete characteristics with the weights found by linear regression.	50

Table 7.1	Characteristics of TCP/UDP/ICMP/IPv4 traffic. Methods used for modeling and comparison are count and ratio (for single value), discrete, and sorted, ordered, and non-keyed sorted for the continuous characteristics.	63
Table 7.2	The goal and calculated similarities for the first five baselines using all characteristics.	68
Table 8.1	The scaled similarities of all the training pairs for all the baselines. These similarities should correspond with the goals in Table J.2.	82
Table 8.2	The new goals and corresponding scaled similarities of all the training pairs for all the baselines. The second set of scaled similarities is produced by dropping the baselines with a goal near (but not equal to) zero from the training set; they are still included in the validation, shown here. . .	87
Table 8.3	The normalized similarities with the scaled similarities for the three weights corresponding to the three different sets of goal values (all values rounded to five significant digits) for all test pairs of real data.	88
Table 9.1	Characteristics which were not assigned weights by linear regression using the scaled goals	94
Table D.1	Count characteristics and Similarity for Baseline 1. Similarity Goal was 1.0 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.	127
Table D.2	Count characteristics and Similarity for Baseline 2. Similarity Goal was 1.0 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.	128
Table D.3	Count characteristics and Similarity for Baseline 3. Similarity Goal was 0.9 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.	128

Table D.4	Count characteristics and Similarity for Basecase 4. Similarity Goal was 0.8 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.	128
Table D.5	Count characteristics and Similarity for Basecase 5. Similarity Goal was 0.1 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.	129
Table D.6	The factors used for the non-linear scaling of the normalized similarities to produce the scaled similarities. The scaled similarity was defined as $bx + cx^2$, where x was the normalized similarity.	131
Table D.7	The goal similarity and composite scaled similarities for the first five base cases. The composite scaled similarity for each basecase was determined by taking the mean of the scaled similarities for all of the measures.	131
Table D.8	The factors used for the non-linear scaling of the normalized similarities to produce the scaled similarities, based on the new goals. The scaled similarity was defined as $bx + cx^2$, where x was the normalized similarity.	131
Table D.9	The new goal similarity and composite scaled similarities for the first five base cases. The composite scaled similarity for each basecase was determined by taking the mean of the scaled similarities for all of the measures.	132
Table D.10	The new goal similarity and composite scaled similarities for the first five base cases, without the “bytes out” characteristic. The composite scaled similarity for each basecase was determined by taking the mean of the scaled similarities for all of the measures.	133
Table D.11	The new goal similarity and composite scaled similarities for the first five base cases, without the “bytes out” or “bytes in” characteristics. The composite scaled similarity for each basecase was determined by taking the mean of the scaled similarities for all of the measures.	133
Table D.12	The normalized similarities of discrete characteristics for the first run of <code>normCompare.pl</code> on basecase 1.	135
Table D.13	The normalized similarities of count and discrete characteristics for basecases 1 through 5, using <code>mapHosts.pl</code>	137

Table D.14	The factors for the non-linear weighting of the count and discrete characteristics, using the goal similarities of (1.0, 1.0, 0.9, 0.8, and 0.1).	138
Table D.15	The factors for the non-linear weighting of the count and discrete characteristics, using the goal similarities of (1.0, 1.0, 0.8465, 0.7283, 0.0394). . .	139
Table D.16	The factors for the linear weighting of the count and discrete characteristics, using the goal similarities of (1.0, 1.0, 0.9, 0.8, 0.1).	140
Table D.17	The expected and calculated similarities for the first five basecases using the count and discrete characteristics and the first linear weighting formula.	140
Table D.18	The new factors (using the fixed formula) for the linear weighting of the count and discrete characteristics, using the goal similarities of (1.0, 1.0, 0.9, 0.8, 0.1).	141
Table D.19	The expected and calculated similarities for the first five basecases using the count and discrete characteristics and the fixed linear weighting formula.	141
Table E.1	The weighting factors for the linear weighting of the singular and discrete characteristics, using the goal similarities of {1.0, 1.0, 0.9, 0.8, 0.1}.	144
Table E.2	The goal and calculated similarities for the first five basecases using the singular and discrete characteristics.	144
Table E.3	The goal and calculated similarities for the first five basecases using the singular and discrete characteristics with the weights found by linear regression.	146
Table H.1	The normalized similarities of all characteristics for basecases 1 through 5.	246
Table H.2	The weighting factors for all characteristics found with linear regression, using the goal similarities of {1.0, 1.0, 0.9, 0.8, 0.1}.	251
Table J.1	Traces used for each training pair for each basecase, along with the intended goal and calculated mean similarity of all normalized characteristics.	287
Table J.2	Scaled, linear, and linear with zeros goals for all training pairs of real data basecases.	291
Table J.3	Weights calculated by linear regression with the scaled goals.	292
Table J.4	Both our calculated scaled similarity (rounded to five significant digits) and Weka's scaled similarity (rounded to three significant digits) for all test pairs of real data, using the scaled training goals.	294
Table K.1	Packets in counts for all traces.	296

Table K.2	Packets out counts for all traces.	299
Table K.3	Connections in counts for all traces.	302
Table K.4	Connections out counts for all traces.	305
Table K.5	Bytes in counts for all traces.	308
Table K.6	Bytes out counts for all traces.	311
Table K.7	SYN-ONLY rate ratios for all traces.	314
Table K.8	SYN-ACK rate ratios for all traces.	316
Table K.9	Idle connection rate ratios for all traces.	319
Table K.10	Half-open connection rate ratios for all traces.	322
Table L.1	Similarity values for individual metrics of basecase 1, train pair 1	418
Table L.2	Similarity values for individual metrics of basecase 1, train pair 2	422
Table L.3	Similarity values for individual metrics of basecase 1, train pair 3	425
Table L.4	Similarity values for individual metrics of basecase 1, test pair 1	429
Table L.5	Similarity values for individual metrics of basecase 1, test pair 2	432
Table L.6	Similarity values for individual metrics of basecase 1, test pair 3	436
Table L.7	Similarity values for individual metrics of basecase 2, train pair 1	439
Table L.8	Similarity values for individual metrics of basecase 2, train pair 2	443
Table L.9	Similarity values for individual metrics of basecase 2, train pair 3	446
Table L.10	Similarity values for individual metrics of basecase 2, test pair 1	450
Table L.11	Similarity values for individual metrics of basecase 2, test pair 3	453
Table L.12	Similarity values for individual metrics of basecase 3, train pair 1	457
Table L.13	Similarity values for individual metrics of basecase 3, train pair 2	460
Table L.14	Similarity values for individual metrics of basecase 3, train pair 3	464
Table L.15	Similarity values for individual metrics of basecase 3, test pair 1	467
Table L.16	Similarity values for individual metrics of basecase 3, test pair 2	471
Table L.17	Similarity values for individual metrics of basecase 3, test pair 3	474
Table L.18	Similarity values for individual metrics of basecase 4, train pair 1	478
Table L.19	Similarity values for individual metrics of basecase 4, train pair 2	481
Table L.20	Similarity values for individual metrics of basecase 4, train pair 3	485
Table L.21	Similarity values for individual metrics of basecase 4, test pair 1	488
Table L.22	Similarity values for individual metrics of basecase 4, test pair 2	492
Table L.23	Similarity values for individual metrics of basecase 5, train pair 1	495
Table L.24	Similarity values for individual metrics of basecase 5, train pair 2	499
Table L.25	Similarity values for individual metrics of basecase 5, train pair 3	502
Table L.26	Similarity values for individual metrics of basecase 5, test pair 1	506

Table L.27	Similarity values for individual metrics of basecase 5, test pair 2	509
Table L.28	Similarity values for individual metrics of basecase 5, test pair 3	513
Table L.29	Similarity values for individual metrics of basecase 6, train pair 1	516
Table L.30	Similarity values for individual metrics of basecase 6, train pair 2	520
Table L.31	Similarity values for individual metrics of basecase 6, train pair 3	523
Table L.32	Similarity values for individual metrics of basecase 6, test pair 1	527
Table L.33	Similarity values for individual metrics of basecase 6, test pair 2	530
Table L.34	Similarity values for individual metrics of basecase 6, test pair 3	534
Table L.35	Similarity values for individual metrics of basecase 7, train pair 1	537
Table L.36	Similarity values for individual metrics of basecase 7, train pair 2	541
Table L.37	Similarity values for individual metrics of basecase 7, train pair 3	544
Table L.38	Similarity values for individual metrics of basecase 7, test pair 1	548
Table L.39	Similarity values for individual metrics of basecase 7, test pair 2	551
Table L.40	Similarity values for individual metrics of basecase 7, test pair 3	555
Table L.41	Similarity values for individual metrics of basecase 8, train pair 1	558
Table L.42	Similarity values for individual metrics of basecase 8, train pair 2	562
Table L.43	Similarity values for individual metrics of basecase 8, train pair 3	565
Table L.44	Similarity values for individual metrics of basecase 8, test pair 1	569
Table L.45	Similarity values for individual metrics of basecase 8, test pair 2	572
Table L.46	Similarity values for individual metrics of basecase 8, test pair 3	575
Table L.47	Similarity values for individual metrics of basecase 9, train pair 1	579
Table L.48	Similarity values for individual metrics of basecase 9, train pair 2	583
Table L.49	Similarity values for individual metrics of basecase 9, train pair 3	586
Table L.50	Similarity values for individual metrics of basecase 9, test pair 1	590
Table L.51	Similarity values for individual metrics of basecase 9, test pair 2	593
Table L.52	Similarity values for individual metrics of basecase 9, test pair 3	594
Table L.53	Similarity values for individual metrics of basecase 10, train pair 1	598
Table L.54	Similarity values for individual metrics of basecase 10, test pair 1	601
Table L.55	Similarity values for individual metrics of basecase 11, train pair 1	605
Table L.56	Similarity values for individual metrics of basecase 11, train pair 2	608
Table L.57	Similarity values for individual metrics of basecase 11, train pair 3	612
Table L.58	Similarity values for individual metrics of basecase 11, test pair 1	615
Table L.59	Similarity values for individual metrics of basecase 11, test pair 2	619
Table L.60	Similarity values for individual metrics of basecase 11, test pair 3	623
Table L.61	Similarity values for individual metrics of basecase 12, train pair 1	626

Table L.62	Similarity values for individual metrics of basecase 12, train pair 2 . . .	630
Table L.63	Similarity values for individual metrics of basecase 12, train pair 3 . . .	633
Table L.64	Similarity values for individual metrics of basecase 12, test pair 1 . . .	637
Table L.65	Similarity values for individual metrics of basecase 12, test pair 2 . . .	640
Table L.66	Similarity values for individual metrics of basecase 12, test pair 3 . . .	644
Table M.1	Weights for the real data basecases sorted from highest to lowest by absolute value.	648

LIST OF FIGURES

Figure 5.1	Plot of SYN-ONLY values as continuous characteristic for one hour of sample data.	43
Figure 8.1	Number of connections for each second after the minute for basecase 4, pair 3	81
Figure 8.2	Portion of packets to privileged services in the connections over the last w seconds for basecase 4, pair 3	83
Figure 8.3	Number of connections for each second after the minute for basecase 7, pair 3	84
Figure 8.4	Number of connections for each second after the minute for basecase 12, pair 2	84
Figure 8.5	Number of connections for each second after the minute for basecase 8, pair 1	85
Figure 8.6	Portion of packets to privileged services in the connections over the last w seconds for basecase 8, pair 1	85
Figure 9.1	Plots of Connection time rate plot for unprivileged connections	91
Figure 9.2	Plots of Packet Destination IP	91
Figure 9.3	Plots of Connection time rate plot for FINs	93
Figure 9.4	Plots of InterPacket delta	93
Figure E.1	Results of Weka's Linear Regression on our normalized similarities from Table 5.1, with the goal values of $\{1.0, 1.0, 0.9, 0.8, 0.1\}$	145
Figure F.1	Plot of Interarrival time of packets values for one hour of sample data.	148
Figure F.2	Plot of Interarrival time of packets values for one hour of sample data with a power function fitted to the data.	148
Figure F.3	Plot of $\log_{10}(\text{Packet Interarrival Time})$ values for one hour of sample data.	150
Figure F.4	Plot of $\log_{10}(\text{Packet Interarrival Time})$ values for one hour of sample data with fitted sixth-degree polynomial.	150
Figure F.5	First piece of fit curve: third-degree polynomial covering about 45% of the data.	151
Figure F.6	Second piece of fit curve: third-degree polynomial covering about 30% of the data. Note overlap with first piece.	151

Figure F.7	Third piece of fit curve: third-degree polynomial covering about 35% of the data. Note overlap with second piece.	152
Figure F.8	Piece-wise function fitted to \log_{10} of packet deltas.	152
Figure F.9	Generated packet interarrival times using \log_{10} values seen in figure Figure F.8.	153
Figure F.10	Plot of interarrival times of packets for both one hour of sample data (green line) and DSL data (blue line). The x-axis of each dataset was scaled separately so that both would fit in the range [0:1]. The y-axis of both datasets were scaled together by the maximum value in both (296.862484) so that both would fit in the range [0:1] while maintaining their relative differences.	153
Figure F.11	Plot of interarrival times of packets for both one hour of sample data (green line) and DSL data (blue line). The x-axis of each dataset was scaled separately so that both would fit in the range [0:1]. The y-axis of each datasets was scaled separately (63.280312 for the blue line and 296.862484 for the green line) so that both would scale the full range of [0:1].	155
Figure F.12	Plot of the \log_{10} of the interarrival times of packets for both one hour of sample data (green line) and DSL data (blue line). The x-axis of each dataset was scaled separately so that both would fit in the range [0:1].	155
Figure F.13	Plot of the \log_{10} of the interarrival times of packets for both one hour of sample data (green line) and DSL data (blue line). The x-axis of each dataset was scaled separately so that both would fit in the range [0:1]. The y-axes of both datasets were scaled by the same amount so that they would both fit in the range [0:1].	157
Figure F.14	Plot of the similarity values (purple line) between the interarrival times of packets for one hour of sample data (green line) and DSL data (blue line). The x-axis of the similarity values aligns with the scaling of the two datasets to both fit in the range [0:1]. The y-axis of the similarity values is true. The y-axes of the two datasets are shown as scaled \log_{10} so that the similarities and differences between them are more visible.	157
Figure G.1	Plot of time to live (TTL) values for one hour of sample data.	160
Figure G.2	Plot of time to live (TTL) values for approx 2.5 days of DSL data.	160

Figure G.3	Plot of original time to live (TTL) values 40–59 (above) and plot of windowed average time to live (TTL) values 38–61 (below), both with a fitted sixth-order polynomial	163
Figure G.4	Inverted FFT with 11 (upper-left) to 4 (lower-right) spectral (complex) values of original TTL values 40–59 – same as original values	165
Figure G.5	Inverted FFT with (from upper-left to lower-right) 13, 12, 11, 9, 8, 6, 5, and 4 spectral (complex) values of windowed average TTL values 40–59; the plots with 10 and 7 spectral values have been omitted for space, and because they look almost identical to the plots with 11 and 8 spectral values, respectively	166
Figure G.6	Illustration of how raw data can be used to find difference in area under the curve	168
Figure G.7	Illustration of how raw data can be used to find difference in area under the curve	168
Figure H.1	Plot of Packets per seconds after the minute values for one hour of sample data.	172
Figure H.2	Plot of Packets per minutes after the hour values for one hour of sample data.	172
Figure H.3	Plot of Packets per hours after midnight UTC values for one hour of sample data.	174
Figure H.4	Plot of Packets per hours after midnight Local values for one hour of sample data.	174
Figure H.5	Plot of Packets per day of the week UTC values for one hour of sample data.	175
Figure H.6	Plot of Bytes transferred per seconds after the minute values for one hour of sample data.	175
Figure H.7	Plot of Bytes transferred per minutes after the hour values for one hour of sample data.	177
Figure H.8	Plot of Bytes transferred per hours after midnight UTC values for one hour of sample data.	177
Figure H.9	Plot of Bytes transferred per hours after midnight Local values for one hour of sample data.	178
Figure H.10	Plot of Bytes transferred per day of the week UTC values for one hour of sample data.	178
Figure H.11	Plot of Packet sizes values for one hour of sample data.	180

Figure H.12	Plot of number of packets over the past w seconds for one hour of sample data.	180
Figure H.13	Plot of number of packets to privileged services versus the number of packets over the past w seconds for one hour of sample data.	183
Figure H.14	Plot of number of packets to unprivileged services versus the number of packets over the past w seconds for one hour of sample data.	183
Figure H.15	Plot of ratio between connections versus packets over the past w seconds for one hour of sample data.	185
Figure H.16	Plot of ratio of privileged connections to all connections over the past w seconds for one hour of sample data.	185
Figure H.17	Plot of ratio of unprivileged connections to all connections over the past w seconds for one hour of sample data.	187
Figure H.18	Plot of ratio of privileged connections to privileged packets over the past w seconds for one hour of sample data.	187
Figure H.19	Plot of ratio of unprivileged connections to unprivileged packets over the past w seconds for one hour of sample data.	188
Figure H.20	Plot of SYN flags versus connections over the past w seconds for one hour of sample data.	188
Figure H.21	Plot of RST flags versus connections over the past w seconds for one hour of sample data.	191
Figure H.22	Plot of FIN flags versus number of connections over the past w seconds for one hour of sample data.	191
Figure H.23	Plot of number of connections versus PSH flags over the past w seconds for one hour of sample data.	192
Figure H.24	Plot of establishment errors versus active connections over the past w seconds for one hour of sample data.	192
Figure H.25	Plot of other errors versus active connections over the past w seconds for one hour of sample data.	194
Figure H.26	Plot of disconnection errors versus active connections over the past w seconds for one hour of sample data.	194
Figure H.27	Plot of average duration of connections active in the past w seconds for one hour of sample data.	196
Figure H.28	Plot of rate of privileged packets out of the past n packets for one hour of sample data.	196

Figure H.29 Plot of rate of unprivileged packets out of the past n packets for one hour of sample data.	198
Figure H.30 Plot of Interarrival time of connections values for one hour of sample data.	198
Figure H.31 Plot of Connections per seconds after the minute values for one hour of sample data.	200
Figure H.32 Plot of Connections per minutes after the hour values for one hour of sample data.	200
Figure H.33 Plot of Connections per hours after midnight UTC values for one hour of sample data.	201
Figure H.34 Plot of Connections per hours after midnight Local values for one hour of sample data.	201
Figure H.35 Plot of Connections per day of the week UTC values for one hour of sample data.	203
Figure H.36 Plot of connection source port values for one hour of sample data.	203
Figure H.37 Plot of connections versus number of packets, over the past m connections, for one hour of sample data.	205
Figure H.38 Plot of privileged connections versus connections, over the past m connections, for one hour of sample data.	205
Figure H.39 Plot of unprivileged connections versus connections, over the past m connections, for one hour of sample data.	207
Figure H.40 Plot of privileged connections versus privileged packets, over the past m connections, for one hour of sample data.	207
Figure H.41 Plot of unprivileged connections versus unprivileged packets, over the past m connections, for one hour of sample data.	208
Figure H.42 Plot of SYN flags versus the number of connections, over the past m connections, for one hour of sample data.	208
Figure H.43 Plot of RST flags versus the number of connections, over the past m connections, for one hour of sample data.	211
Figure H.44 Plot of FIN flags versus the number of connections, over the past m connections, for one hour of sample data.	211
Figure H.45 Plot of number of connections versus the number of PSH flags, over the past m connections, for one hour of sample data.	212
Figure H.46 Plot of number of establishment errors versus the number of connections, over the past m connections, for one hour of sample data.	212

Figure H.47 Plot of number of other errors versus the number of connections, over the past m connections, for one hour of sample data.	214
Figure H.48 Plot of disconnection errors versus the number of connections, over the past m connections, for one hour of sample data.	214
Figure H.49 Plot of average duration, over the past m connections, for one hour of sample data.	216
Figure H.50 Plot of packets per connection values for one hour of sample data. . . .	216
Figure H.51 Plot of packets sent per connection values for one hour of sample data. . . .	217
Figure H.52 Plot of packets received per connection values for one hour of sample data. . . .	217
Figure H.53 Plot of connection duration values for one hour of sample data.	219
Figure H.54 Plot of connection percent control packets values for one hour of sample data.	219
Figure H.55 Plot of connection percent data packets values for one hour of sample data.	221
Figure H.56 Plot of bytes per connection values for one hour of sample data.	221
Figure H.57 Plot of bytes sent per connection values for one hour of sample data. . . .	223
Figure H.58 Plot of bytes received per connection values for one hour of sample data. . . .	223
Figure H.59 Plot of data bytes per connection values for one hour of sample data. . . .	224
Figure H.60 Plot of data bytes sent per connection values for one hour of sample data. . . .	224
Figure H.61 Plot of data bytes received per connection values for one hour of sample data.	226
Figure H.62 Plot of connection frag packet rate values for one hour of sample data. . . .	226
Figure H.63 Plot of connection wrong frag packet rate values for one hour of sample data.	228
Figure H.64 Plot of connection source window size values for one hour of sample data.	228
Figure H.65 Plot of connection source window size values for approx 2.5 days of DSL data.	230
Figure H.66 Plot of connection destination window size values for one hour of sample data.	230
Figure H.67 Plot of connection urgent packet rate values for one hour of sample data. . . .	231
Figure H.68 Plot of connection resend packet rate values for one hour of sample data. . . .	231
Figure H.69 Plot of connection wrong resend packet rate values for one hour of sample data.	233
Figure H.70 Plot of connection duplicate ACK packet rate values for one hour of sample data.	233

Figure H.71	Plot of wrong ACK errors per connection values for one hour of sample data.	235
Figure H.72	Plot of connection wrong data packet size rate values for one hour of sample data.	235
Figure H.73	Plot of window exceeded errors per connection values for one hour of sample data.	236
Figure H.74	Plot of connection hole rate values for one hour of sample data.	236
Figure H.75	Plot of connection errors values for one hour of sample data.	238
Figure H.76	Plot of reset connection values for one hour of sample data.	238
Figure H.77	Plot of other errors per connection values for one hour of sample data.	240
Figure H.78	Plot of disconnection errors values for one hour of sample data.	240
Figure H.79	Plot of number of connections per source IP for one hour of sample data.	242
Figure H.80	Plot of number of connections per destination IP for one hour of sample data.	242
Figure H.81	Plot of number of packets per source IP for one hour of sample data.	244
Figure H.82	Plot of number of packets per destination IP for one hour of sample data.	244
Figure H.83	Plot of number of bytes transferred per source IP for one hour of sample data.	245
Figure H.84	Plot of number of bytes transferred per destination IP for one hour of sample data.	245
Figure K.1	Packet Service distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	326
Figure K.2	Bytes Service distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	327
Figure K.3	Connection Service distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	328
Figure K.4	Packet Source port distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	329
Figure K.5	Bytes Source port distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	330
Figure K.6	Connection Source port distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	331

Figure K.7	Connection Source port distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	332
Figure K.8	Packet TTL distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	333
Figure K.9	Packet TTL distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	334
Figure K.10	InterPacket delta distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	335
Figure K.11	Packet sec distributions for all traces, with all the traces from the same dataset plotted together.	336
Figure K.12	Packet min distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	337
Figure K.13	Packet GmHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	338
Figure K.14	Packet LocHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	339
Figure K.15	Packet weekday distributions for all traces, with all the traces from the same dataset plotted together.	340
Figure K.16	Bytes sec distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	341
Figure K.17	Bytes min distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	342
Figure K.18	Bytes GmHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	343
Figure K.19	Bytes LocHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	344
Figure K.20	Bytes weekday distributions for all traces, with all the traces from the same dataset plotted together.	345
Figure K.21	Packet size distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	346
Figure K.22	Packets in last w secs distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	347

Figure K.23	Priv packets time rate distributions for all traces, with all the traces from the same dataset plotted together.	348
Figure K.24	Unpriv packets time rate distributions for all traces, with all the traces from the same dataset plotted together.	349
Figure K.25	Connections time rate distributions for all traces, with all the traces from the same dataset plotted together.	350
Figure K.26	Priv connections connection time rate distributions for all traces, with all the traces from the same dataset plotted together.	351
Figure K.27	Unpriv connections connection time rate distributions for all traces, with all the traces from the same dataset plotted together.	352
Figure K.28	Priv packets priv connection time rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	353
Figure K.29	Unpriv packets unpriv connection time rate distributions for all traces, with all the traces from the same dataset plotted together.	354
Figure K.30	SYNs connection time rate distributions for all traces, with all the traces from the same dataset plotted together.	355
Figure K.31	RSTs connection time rate distributions for all traces, with all the traces from the same dataset plotted together.	356
Figure K.32	FINs connection time rate distributions for all traces, with all the traces from the same dataset plotted together.	357
Figure K.33	PSH connection time rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	358
Figure K.34	Establishment errors connection time rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	359
Figure K.35	Other errors connection time rate distributions for all traces, with all the traces from the same dataset plotted together.	360
Figure K.36	Disconnection errors connection time rate distributions for all traces, with all the traces from the same dataset plotted together.	361
Figure K.37	Ave duration over last w secs distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	362

Figure K.38 Priv packets packet rate distributions for all traces, with all the traces from the same dataset plotted together.	363
Figure K.39 Unpriv packets packet rate distributions for all traces, with all the traces from the same dataset plotted together.	364
Figure K.40 InterConnection delta distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	365
Figure K.41 Connection sec distributions for all traces, with all the traces from the same dataset plotted together.	366
Figure K.42 Connection min distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	367
Figure K.43 Connection GmHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	368
Figure K.44 Connection LocHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	369
Figure K.45 Connection weekday distributions for all traces, with all the traces from the same dataset plotted together.	370
Figure K.46 Connection packet rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	371
Figure K.47 Connection Priv connections rate distributions for all traces, with all the traces from the same dataset plotted together.	372
Figure K.48 Connection Unpriv connections rate distributions for all traces, with all the traces from the same dataset plotted together.	373
Figure K.49 Connection Priv packet rate distributions for all traces, with all the traces from the same dataset plotted together.	374
Figure K.50 Connection Unpriv packet rate distributions for all traces, with all the traces from the same dataset plotted together.	375
Figure K.51 Connection SYNs rate distributions for all traces, with all the traces from the same dataset plotted together.	376
Figure K.52 Connection RSTs rate distributions for all traces, with all the traces from the same dataset plotted together.	377

Figure K.53 Connection FINs rate distributions for all traces, with all the traces from the same dataset plotted together.	378
Figure K.54 Connection PSH rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	379
Figure K.55 Connection Establishment errors rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	380
Figure K.56 Connection Other errors rate distributions for all traces, with all the traces from the same dataset plotted together.	381
Figure K.57 Connection Disconnection errors rate distributions for all traces, with all the traces from the same dataset plotted together.	382
Figure K.58 Ave duration over last m connections distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	383
Figure K.59 Number of packets distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	384
Figure K.60 Number of packets in distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	385
Figure K.61 Number of packets out distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	386
Figure K.62 Duration distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	387
Figure K.63 Number control packets rate distributions for all traces, with all the traces from the same dataset plotted together.	388
Figure K.64 Number data packets rate distributions for all traces, with all the traces from the same dataset plotted together.	389
Figure K.65 Number bytes transferred distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	390
Figure K.66 Number bytes transferred in distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	391

Figure K.67 Number bytes transferred out distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	392
Figure K.68 Number data bytes transferred distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	393
Figure K.69 Number data bytes transferred in distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	394
Figure K.70 Number data bytes transferred out distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	395
Figure K.71 Fragmented packets rate distributions for all traces, with all the traces from the same dataset plotted together.	396
Figure K.72 Bad fragment rate distributions for all traces, with all the traces from the same dataset plotted together.	397
Figure K.73 Max Src Window distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	398
Figure K.74 Max Dst Window distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	399
Figure K.75 Urgent rate distributions for all traces, with all the traces from the same dataset plotted together.	400
Figure K.76 Resend rate distributions for all traces, with all the traces from the same dataset plotted together.	401
Figure K.77 Wrong resend rate distributions for all traces, with all the traces from the same dataset plotted together.	402
Figure K.78 Duplicate ACK rate distributions for all traces, with all the traces from the same dataset plotted together.	403
Figure K.79 Wrong ACK distributions for all traces, with all the traces from the same dataset plotted together.	404
Figure K.80 Wrong data packet size rate distributions for all traces, with all the traces from the same dataset plotted together.	405
Figure K.81 Window exceeded rate distributions for all traces, with all the traces from the same dataset plotted together.	406

Figure K.82 Hole rate distributions for all traces, with all the traces from the same dataset plotted together.	407
Figure K.83 Number connection errors distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	408
Figure K.84 Number reset connection distributions for all traces, with all the traces from the same dataset plotted together.	409
Figure K.85 Number other errors distributions for all traces, with all the traces from the same dataset plotted together.	410
Figure K.86 Number disconnection errors distributions for all traces, with all the traces from the same dataset plotted together.	411
Figure K.87 Packet Destination IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	412
Figure K.88 Bytes Destination IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	413
Figure K.89 Connection Destination IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	414
Figure K.90 Packet Source IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	415
Figure K.91 Bytes Source IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	416
Figure K.92 Connection Source IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.	417
 Figure M.1 Plots of characteristics 1 through 8 from basecase 3, pair 3	666
Figure M.2 Plots of characteristics 9 through 16 from basecase 3, pair 3	667
Figure M.3 Plots of characteristics 17 through 24 from basecase 3, pair 3	668
Figure M.4 Plots of characteristics 25 through 32 from basecase 3, pair 3	669
Figure M.5 Plots of characteristics 33 through 40 from basecase 3, pair 3	670
Figure M.6 Plots of characteristics 41 through 48 from basecase 3, pair 3	671
Figure M.7 Plots of characteristics 49 through 56 from basecase 3, pair 3	672
Figure M.8 Plots of characteristics 57 through 64 from basecase 3, pair 3	673

Figure M.9	Plots of characteristics 65 through 72 from basecase 3, pair 3	674
Figure M.10	Plots of characteristics 73 through 80 from basecase 3, pair 3	675
Figure M.11	Plots of characteristics 81 through 88 from basecase 3, pair 3	676
Figure M.12	Plots of characteristics 89 through 92 from basecase 3, pair 3	677
Figure M.13	Plots of characteristics 1 through 8 from basecase 8, pair 3	678
Figure M.14	Plots of characteristics 9 through 16 from basecase 8, pair 3	679
Figure M.15	Plots of characteristics 17 through 24 from basecase 8, pair 3	680
Figure M.16	Plots of characteristics 25 through 32 from basecase 8, pair 3	681
Figure M.17	Plots of characteristics 33 through 40 from basecase 8, pair 3	682
Figure M.18	Plots of characteristics 41 through 48 from basecase 8, pair 3	683
Figure M.19	Plots of characteristics 49 through 56 from basecase 8, pair 3	684
Figure M.20	Plots of characteristics 57 through 64 from basecase 8, pair 3	685
Figure M.21	Plots of characteristics 65 through 72 from basecase 8, pair 3	686
Figure M.22	Plots of characteristics 73 through 80 from basecase 8, pair 3	687
Figure M.23	Plots of characteristics 81 through 88 from basecase 8, pair 3	688
Figure M.24	Plots of characteristics 89 through 92 from basecase 8, pair 3	689

PREFACE

This work was originally supposed to be “The Application of Data Mining to Network Intrusion Detection”. That is, in fact, the subject and title of the dissertation proposal put before the author’s qualification committee. The committee complimented the author on the thoroughness of the survey of prior work, and proceeded to lob a single volley with the question, “How are you going to test your system?”

The author was rather nonplussed, as the proposal clearly outlined that the work would utilize the DARPA Intrusion Detection Evaluation dataset for testing; it noted that while there were known issues with it, Mahoney and Chan’s work indicated that if advanced intrusion detection system could not perform well against the DARPA data, it wouldn’t stand a chance in the real-world, hence it should provide a good baseline assessment of the system; furthermore, the use of a standard dataset in testing allowed for direct comparisons of systems. The committee was unmoved.

Over the next several months, while waiting for the chair to come back with a list of necessary changes, the author proceeded with the proposed research, starting with a baseline assessment of the DARPA data using Snort. The key finding here was that while the attacks in the DARPA data may have been modeled correctly, *and* demonstrate attacks that signature-based IDS have trouble detecting, there was nothing to demonstrate that the background traffic had any similarity to real network traffic. In other words, the DARPA data can be used to evaluate the true positive response of a system, but not the false positive response.

The author now understood the committee’s point. In order for not just his research – but all scientifically sound network intrusion detection research – to continue, we would need a new dataset. The generation of a dataset which would look sufficiently similar to real network traffic for NIDS research sounded like a dissertation topic in itself, so the author set down a path for a new proposal.

About a month into researching traffic generation techniques, ideas started to gel, and the author asked, “If the key criticism of the DARPA data is that it wasn’t validated against real network traffic, how do I do that for the traffic I’m generating?” Given how casually the criticism was made by McHugh, surely there must be an accepted method to perform such validation. A search of the literature revealed that the network engineering community has extensively compared bandwidth and latency, and measured some other metrics, but no one had done a holistic comparison of all the aspects of network traffic that we care about when doing advanced intrusion detection.

That was the genesis of this work.

ACKNOWLEDGMENTS

Before going into all my personal acknowledgments, I must say that the most difficult thing about this research area is obtaining real network data, as even with modern anonymization techniques, most organizations are reluctant to release any network trace data – even to their own research staff (be they a student or an employee). With this in mind, I must start by acknowledging all the data repositories I used for this work:

- We gratefully acknowledge the use of network data from the CRAWDAD archive at Dartmouth College (Dartmouth College 2008).
- We gratefully acknowledge the use of network data from the DatCat archive at CAIDA (Cooperative Association for Internet Data Analysis (CAIDA) 2008).
- We gratefully acknowledge the use of DefCon CCTF data from Shmoo (Shmoo 2007).

And a huge thank you to the organizations who provided data to these archives!

The following work was eight years in the making, representing a quarter of my life. I don't think I'll take on another project as consuming as this one is, so this is my best chance to thank everyone that helped me not only on this project specifically, but who got me to this point.

First: A big thank you to my parents, Ronald Brugger – who taught me (among other things) you don't need to know much about computers to be a computer security expert – and Susan Brugger – who does know a lot about computers and introduced me to the field.

All my friends, particularly the ones who were there for me in good times and bad, in rough chronological order: Sean, Brian, Aaron, Dan, Stacy, Darren, and Shirley. A real big thanks to Tammy and Kevin Roust, John and Jenn Brown, and James Schek and Candace Bolles, who all whom were not only great friends, but who tried to help me collect data for this work, even if it didn't pan out.

Thanks also to David Steich and everyone on the TIGER team for showing me the value of the PhD mind set, and Kim Mish who convinced me to go straight for the PhD, and that Davis was the place to do it.

This work has its origin in the CIAC Logger project; thanks to the original team there for the motivating ideas: Marcey Kelley, Ken Sumikawa, and Shaun Wakumoto.

While Lawrence Livermore National Laboratory didn't pay for any of the research herein, I must acknowledge that it paid most of the fees and tuition for my graduate studies. Many thanks to Kathy Zobel in Employee Development and her compatriot-in-education Claire Daughtry of the ITV program at Davis, without whom I couldn't have gone to grad school and

still paid a mortgage. The Lab provided countless cheerleaders and many mentors, chief among these was Tina Eliassi-Rad, who was always willing to provide advice at all hours, any day of the week.

All my teachers over the years – okay, most all of them – I might not remember all your names, but you took me through the steps to make it here. In particular, Richard Meier, for instilling the joy of formal education in me; Mr Simonson, whose nickname for me was “Dr. T”; Vic Rames, who drilled the scientific into me; and Dr Stan Pickard, who mentored, opened doors, and showed me whole new worlds. Also a big thank you to Mary Homan – you can add me to the card catalog now!

Most all of my undergraduate CS profs were great. In particular, my undergrad CS research adviser, Vernon Rego, took me under his wing when I was still green, and I hope he’ll forgive me that my future was at Davis and not his lab. Thanks also to Gene Spafford whom I never had the honor of taking a class from, but whose weekly security seminars got me interested in the field. The first one I saw was from Terran Lane, which planted the seed that led to the above mentioned Logger and my original dissertation proposal. Finally, thanks to Aimée Surprenant, my undergrad Psych research adviser for introducing me to the joys of graduate studies.

Obviously, my graduate professors have been most influential on this work. A huge thank you to Wenke Lee, whose dissertation formed the foundation my original proposal was based off of, who showed me what a dissertation proposal should look like, who was very supportive in my efforts to collect data, and who (after myself) traveled further for my research than anyone else. Michael Gertz and Rao Vemuri were both great in the classroom, and generously served on my qualification examination committee. I hope I hold the record for distance traveled (California to D.C.) to track down Karl Levitt, but that’s only because his expertise to critically evaluate the new proposal that led to this work was that valuable, and I appreciate his making the time to see me. I’m not sure Felix Wu and Prasant Mohapatra knew what they were agreeing to when they agreed to be on my dissertation committee, so a big thank you to both of you! Finally, this work would not have been possible without one more member of my quals committee, the chair of my dissertation committee, graduate adviser, mentor, and fellow Boilermaker, Matt Bishop. He put in a lot of effort for nothing more than the call of good science and an incomparable debt of gratitude (and maybe some copy editing and a journal review).

This work is dedicated to the Kara Brugger Home for Wayward Computer Scientists, run by my loving wife and staffed by Furrball, Fuzzbucket, and Fuufbrain who stayed out of my way more often than not, and were always there when I needed them.

1

Introduction

Network science is in its infancy. While computer networks continue to permeate every aspect of modern living, we have only scratched the surface of studying the emergent behaviors of these systems. In particular, we have no accepted means of comparing two network traces and determining – quantitatively – how similar they are. Having the means to do so would significantly advance research in the areas of network forensics, administration, application and protocol development, traffic generation, and intelligence. This work proposes a methodology to compare two computer network traces and assign a quantitative similarity value to those traces, and tests an instantiation of this methodology on IPv4 traffic.

The key contributions of this work are a list of characteristics of IPv4 traffic which can be used for such a comparison, methods to extract those characteristics from network traces, a set of methods for modeling these characteristics, and an associated set of methods for comparing them.

The remainder of this chapter states the problem we examine in the rest of the work, followed by our general approach, an outline of the rest of the work, and a summary of our key contributions.

1.1 Problem statement

One of the key criticisms leveraged by McHugh (2000) towards the artificial network traces produced for the DARPA IDS Evaluation was that they had never been validated to look like real network traffic. Unfortunately, there is no accepted means of comparing network traces to validate that they look sufficiently similar. Currently, any such comparison is done as a qualitative assessment by someone well versed in network protocols, or is done on a small

subset of characteristics of the trace, such as throughput and interpacket latency. While each approach has sufficed for particular purposes – such as network engineering – by not considering the characteristics of the network traces holistically, the resulting assessments are not generally applicable.

Our goal in this work is to provide a methodology for quantitative network comparison. Given an instance of this methodology, two people should be able to compare and contrast any two network traces in quantitative terms, including an overall similarity value. In order to do this, we need a solid definition of similarity. Most of the informal work on network similarity to date has used a general definition of *similarity* such as, “a Gestalt principle of organization holding that (other things being equal) parts of a stimulus field that are similar to each other tend to be perceived as belonging together as a unit” (Princeton University Cognitive Science Laboratory 2006). We are interested in a more formal definition of similarity for network traces. We will begin by considering an ideal definition of *network equality*: “Two network traces are equal if a bit-by-bit comparison between the two does not show any difference.” This is not very useful, however, because there is likely information in the trace that we are not interested in. For example, there are numerous “reserved” fields in network headers which may have different values which do not matter to the end applications. As such, we may want to use a definition for network equality such as, “Two network traces are equal if the values of the identified fields for each sequential packet in each trace are equal.” We are much closer now, but we must also consider that we might be interested in the relative value of the fields, as opposed to the absolute values. Instead of simple field values, we will consider the distribution of characteristics:

Definition. Network Equality: Two network traces are equal if the distribution of identified characteristics from each of the traces is the same.

When using a distribution of the actual values in the order they appear, this definition becomes equivalent to the previous one. Given this definition of equality, we can define similarity:

Definition. Network Similarity: Two network traces which are equal shall have a similarity of 1.0, and a non-null network trace compared with a null (empty) network trace shall have a similarity of 0.0. The extent to which two non-null, non-equal network traces deviate from equality shall determine the similarity value in the range (0..1), as determined by the weighted similarities of the distributions of identified characteristics.

In the next section, we provide a mathematical definition of our similarity formula. The similarity value will be explainable based on a report of the measures used to produce it. We expect the overall similarity value to be similar to an overall similarity value an expert would assign to the

two traces based on the size, topology, user base, application mix, and applicable policies of the network at the time each trace was taken.

The applicability of such a value will depend on the domain. For example, when attempting to create artificial network traces which look sufficiently similar to real network traffic, we may specify an overall similarity of > 0.9 , whereas comparing a trace with a new application protocol to one without may only be concerned if the similarity is < 0.75 . We discuss the potential applications of this methodology at greater length in Chapter 2. Beyond the applicability of the absolute values, the values should be useful when compared relatively; that is, two traces with a similarity value of 0.9 are more similar than those with a similarly value of 0.8 and much more similar than those with a value of 0.4.

1.2 Approach

Our approach to provide such a comparison is to extract all the characteristics ($\kappa_1.. \kappa_n$) we can think of from a pair of traces, compute a similarity value between the traces for each of these characteristics $(1 - \frac{|\kappa_{1,1} - \kappa_{1,2}|}{\kappa_{1,1} + \kappa_{1,2}})$, and combine these individual similarity values (using a weighting – ω – for each characteristics) into an overall similarity value for the two traces:

$$\sum_{i=1}^n \omega_i \left(1 - \frac{|\kappa_{i,1} - \kappa_{i,2}|}{\kappa_{i,1} + \kappa_{i,2}} \right)$$

. We do this in two stages; in the first we enumerate the characteristics we can compare and compute the weights we will use to combine the individual similarities together; we can then perform the second stage of using these characteristics and weights to actually compare traces as often as we wish. We explain the process in much more detail in Chapter 3. After that, we demonstrate this approach using an instantiation of it for IPv4 traffic.

1.3 Outline

In the next chapter, we explain why there is a need for quantitative network comparison, including a look at prior work in the area. Given that, we will provide a brief overview of our methodology for quantitatively comparing network traces. Then we start to get into our sample instantiation of the methodology by explaining how we constructed our initial artificial test data, and the tests we did on it using count based and discrete distribution characteristics. Then we take a short, necessary detour to explore the ambiguities with many of the network connection characteristics that we use, and the formal definitions that we used to resolve that ambiguity.

With that, we continue testing with the artificial test data by comparing the continuous characteristics. We then turn to real data by building up our base cases and running our instantiation of the methodology with them. This is followed by an analysis of the performance on the real data. Finally, we present our conclusions.

1.4 Conclusion

This work presents a methodology for quantitatively comparing two network traces, and demonstrates an instantiation of it for IPv4 traffic. The key contributions of this work are a list of characteristics of IPv4 traffic which can be used for such a comparison, a set of methods for modeling these characteristics, and an associated set of methods for comparing them.

2

The Need for Network Comparison

2.1 Usefulness of a More Holistic Comparison

There are numerous applications that would benefit from a more holistic quantitative comparison of two network traces:

1. Network Forensics
2. Network Administration
3. Network Application Development
4. Network Protocol Development
5. Network Traffic Generation
6. Network Intelligence

We will look at each briefly.

2.1.1 Network Forensics

Given two traces of two different attack incidents, a holistic comparison of the traces will allow us to compare and contrast the attacks at a network level. Hopefully, this will allow some insight into the modus operandi of the attacks. This will go beyond a “Network Ballistics”

type comparison (as in (Bartoletti 2004; Parno and Bartoletti 2004), which focuses on the tools used and their settings, to consider the set of tools used for the entire attack, and human factors such as typing latency if a shell connection is made.

2.1.2 Network Administration

Administrators of large networks can have a hard enough time dealing with issues such as network reliability and policy enforcement; frequently they do not have enough time to consider how – and more importantly, why – their network is changing. Some quick measurements from their gateway might tell them that bandwidth usage is up X% this month, but without knowing why, it is difficult to plan appropriately. A method to quantifiably compare a trace of the network from say, a month ago, to a trace today, could tell not only how much the network has changed, but what factors accounted for that change. Perhaps the change is due to a new file sharing protocol that policy dictates should be blocked. Or perhaps the change is due to an increase in video teleconferencing, which must take latency into account as much as bandwidth when planning for future capacity.

2.1.3 Network Application Development

Network applications are constantly growing in importance. For numerous reasons, application developers are moving away from monolithic desktop applications to server based applications, many of which are accessible from standard web browsers. Additionally, the number of applications that retrieve – particularly those that stream – data from the Internet, continues to increase. Many network-centric applications are sensitive to changes in the network they are operating on, and may themselves change the behavior of that network. It is likely that the application developer will not be able to see these changes on their local network that they use for development. Instead, they will need to take a snapshot of the network with and without their application running, in order to compare them to gain an understanding of how their application is affecting the network. Such a comparison tool would be of additional use in allowing the developer to see how changes to the application change the way that it interacts with the network.

2.1.4 Network Protocol Development

Closely related to network application development, is the development of network protocols. This may mean either the development of new protocols or modifications of existing protocols, at anywhere from the network to the application level. One might think of the differ-

ence as application development is concerned with how the use of established protocols impacts the network, whereas in protocol development we are concerned with how changes to the protocol itself will impact the network. This is something that Floyd and Kohler (2003) and Medina et al. (2005) tackle in Floyd and Kohler and Medina et al., respectively. Both papers discuss what measurements are useful when doing network protocol development and how the choice of measurements has impacted protocol development. Our approach, as we will discuss, is to include all possible characteristics and determine relative relevance as one of the steps when quantifying similarity.

2.1.5 Network Traffic Generation

Just as protocol development was closely related to application development, so too is traffic generation closely related to protocol development, primarily because simulation provides a support on which much protocol development is done. Traffic simulation and generation are used for more than just protocol development though: they are used to test all sorts of things that interoperate with the network, spanning the gambit from software applications (such as intrusion detection systems), to hardware (such as routers).

It was actually this area that spawned the author's original interest in network comparison. Specifically, the generation of test data for intrusion detection systems. This is chronicled in (Brugger and Chow 2007).

2.1.6 Network Intelligence

Numerous organizations are increasing interested in understanding what is going on inside various computer networks. We will dub this general category "Network Intelligence". Such work may be interested in identifying the type of traffic traveling over an encrypted connection¹, or perhaps we have a trace of some subnet and we want to see what other subnet it is most similar to for provisioning purposes.

There is also a much deeper, philosophical argument for Network Intelligence, in that understanding how networks work is an end in itself. While computer science at its core is interested in understanding how computers operate, network research at its core must be interested in understanding how networks operate at the network layer and above. Being able to compare two network traces in a quantitative manner is a step in this direction.

¹ssh presents an interesting conundrum to security aware organizations: its use may be encouraged over protocols such as telnet so that adversaries can not see what's being sent over the connection, but it also means that the policy enforcement personnel can not see what the connection is being used for either.

2.2 Previous work

There has been surprisingly little research on network simulation for the purposes of testing network based security systems. Most all of the network modeling and simulation work to date has been done by the network infrastructure community for the purposes of developing and testing more efficient queuing algorithms and new protocols. In this arena, the interest is primarily in flow rates and interarrival times. While these things are certainly important when simulating network traffic for security purposes – for instance to ensure that devices can handle normal network loads – it ignores much of the internals of the traffic, for example the application mix. When things like the application mix are considered, the focus is usually on the few applications that make up the bulk of network traffic flow, ignoring the low traffic applications that may be of considerable importance to security devices (such as RPC and SMB services).

When modeling the network flows for internetworking purposes, relatively statistical metrics have been sufficient to validate the models, and hence the correctness of the simulations created by those models. The advances in network modeling have been driven primarily by the use of better statistical measures of the network flows.

From the early days of computer networks, many assumptions were made about network packet flows, primarily that traffic was dominated by bulk transfer operations, which moved a large quantity of data, that interactive applications produced fairly symmetric flows, and that packets followed a Poisson distribution model at any given point in the network. In 1992, Danzig et al. made a number of discoveries that advanced our understanding and ability to model networks. Specifically:

- Different applications produce different packet interarrival times.
- Interarrival characteristics need to be modeled at the application level since network infrastructure changes can change them.
- Interarrival times are only interesting for interactive applications since they will change for bulk transfer applications with network improvements.
- Bulk data is not necessarily large or unidirectional, and interactive traffic is not symmetric in size in both directions.
- The distribution of network endpoint pairs differs between applications

Danzig et al. measured a large number of traffic characteristics in the course of this research including packet interarrival times and size, session bytes transferred (in each direction and bidirectionally), duration and packets transferred, items per connection for bulk and item sizes,

network pair distribution, probability of concurrent connections to same network or host, connection rate by time of day, and MTU.

Despite the large of parameters that they considered, the work is primarily focused on the interests of the network engineering community. In particular, the authors only considered the data portion of the connection, and ignored packets that did not produce a significant load, such as establishment, tear down, ACK-only packets, retransmitted packets, and anomalous packets. Such packets are important from a security perspective however, and are likely useful to the network protocol community as well.

The following year Leland et al. released the paper (expanded in 1994) that set the stage for network research for the next decade by showing that Ethernet packet traffic (including traffic introduced from the Internet) had self-similar – or fractal – interarrival times. Specifically, the statistical nature of the interarrival times followed the same long-tail distribution at multiple time scales.

While certainly an interesting result with unquestionable ramifications for router development, this unfortunately caused an almost singular focus on fractal patterns for interarrival times within the network modeling community:

1. Erramilli et al. (1994) showed how chaotic maps could be used to model and produce the self-similar distributions of packet interarrival times.
2. Paxson and Floyd (1995) established that Poisson processes were useful for modeling the interarrival times of user-initiated sessions, but that other session interarrival times, and the packet interarrival times within all WAN sessions, are best modeled using self-similar processes.
3. Park (1997) classifies the research into the causes of self-similarity. Evidence was found for multiple sources, including packet queuing by network stacks, protocol reliability techniques, application characteristics, and the nature of the data itself being transmitted. Park notes that source based simulation is necessary for proper generation of network traffic because direct packet generation fails to capture the interactions between components which creates the fractal pattern in the first place.
4. Sikdar and Vastola (2001) went on to provide a mathematical formalism for why TCP, in particular its retransmission features, can create interpacket arrival times that follow a fractal pattern, even within an individual connection. In particular, they show that the self-similarity of a connection is directly proportional to the loss rate on the connection. They go on to hypothesize that this behavior may vary – however will not vanish – depending on the TCP implementation used.

5. Racz et al. (2003) expanded the work in (Sikdar and Vastola 2001) by looking at the dominate cause of self-similarity under different conditions. They found that synchronization between sessions was the dominate cause in low-congestion and low-loss networks. We assume that this is due to network queuing effects as noted in earlier papers. In the face of modest congestion, the TCP congestion avoidance protocol quickly becomes the dominate contributor to self-similarity. We find it interesting that Racz et al. looked only at the network layer and above, as the original self-similarity paper by Leland et al. considered all Ethernet traffic, hence implying that the self-similarity was an artifact of the data-link layer. Perhaps unknowingly, Leland et al. imply that the Ethernet congestion avoidance protocol is a cause of this property when they suggest that the congestion avoidance protocol for ISDN-B be carefully analyzed to be robust in the face of fractal patterned traffic. We wonder if congestion avoidance at the data link layer was only a significant contributor to self-similarity when Ethernet was used in a shared segment topology, as in Leland et al., as opposed to the switched topology most likely used by Racz et al.. An interesting test would be to repeat Racz et al.'s experiments on a data link protocol such as token ring. In any case, once the network starts having a high loss rate, TCP's resend mechanism takes over for the congestion avoidance protocol as the primary contributor to self-similarity.
6. In a departure from the network engineering community, physicists interested in the nature of scale-free networks have become interested in the Internet. In Yook et al. (2002) the authors show that topology of hosts on the Internet follows a fractal (self-similar) pattern ². While they do not investigate the result this has on the network traffic, one wonders if the self-similar nature of the network topology is one of the causes of self-similar interpacket times for network traffic.
7. Willinger et al. (2002) examines this previous work in the context of a proposed validation model, which examines the premises that the model is based on. Though the use of this validation model, they serve to lend credence to the traffic models of the network engineering community, and reject the basis of the models from the physics community. While their validation model does serve to support or reject traffic models based on their premises, it does not examine the generated traffic itself, nor the characteristics therein.

Floyd and Paxson (2001) followed up their seminal 1995 paper with an excellent examination of why simulating Internet traffic is so hard. The most valuable contribution of this paper is their list of invariants which have held for Internet traffic to date, and which are likely to

²While the authors call it the “physical layout” of the Internet, link-layer protocols such as label-based switching abstract the network layer topology from the physical topology.

continue to hold for the foreseeable future. Fortunately, it was here that we begin to see a break from a singular focus on self-similarity which, while certainly a factor in the list of invariants, is not the sole invariant. The paper also succinctly summarizes the important conclusions of the above research that proceeded it.

Mellia et al. (2002) made a huge push to look at the network traffic as more than just the interarrival time between the sessions and packets, and the related packet size, session duration and bytes transferred, or loss rate. They put together a tool (Tstat) that can produce 80 different graphs or plots of network performance characteristics from a network trace. This was a significant advance to allow an analyst to see the effect of various TCP/IP settings on performance; however the focus was squarely on performance and not security, and it did not provide any means to automatically model the network such that new traces could be produced with the same characteristics. For example, no mention is made of measures of the use of IP options or overlapping IP fragments, either of which may be used to attempt to break certain TCP stack implementations. Further, there does not appear to be any correlation between various metrics and the source or source / dest pair, which would be useful to create a trace with the same characteristics, and identify hosts with conflicting measures, such as a host appearing to use both a 64 and a 128 TTL, which may indicate its attempting to probe the network, or may just be acting as a network address translation gateway.

The most complete system for modeling and validating network traces is presented in (Lan and Heidemann 2002). The authors build application level source models by hand, then parameterize them with distributions automatically built from network traces. The traces produced when the models are used for simulation are then validated using quantitative and qualitative statistical measures. This work pulled together many positive qualities, such as:

- They recognized that the models were the same for different networks, and that it is the parameterization that dictates the unique characteristics.
- They demonstrated the importance of application level parameters (in this case, the distribution of object sizes in HTTP).
- They use multiple quantitative statistical comparisons (Wavelet Scaling Plot and Kolmogorov-Smirnov Goodness of Fit Test) to validate the generated traces.
- For the Kolmogorov-Smirnov test they used the most restrictive critical value of significance so as not to dictate a distribution.
- Instead of just measuring the packet interarrival time, they also used the packet size, session duration, size, and interarrival, and the per-host interarrival and duration, as well as the

protocol mix and traffic volume for comparison.

- Additional parameters were used for the models, specifically, TCP window size, Round Trip Time, and bottleneck bandwidth.

Nevertheless, their work left many questions unanswered:

- While the use of quantifiable statistical measures for comparison was a large leap ahead of the standard Hurst parameter estimation that most work used in the decade that proceeded it, no justification was given for the use of the Kolmogorov-Smirnov test or the Wavelet Scaling Plot (although there is some precedent for the latter in the network engineering community).
- The presented results – particularly that the inbound and outbound traffic from a network, and the same network at two different times, and two different networks differ above the level of significance – are intuitive. Further, it seems convenient that the same statistical measures validate that the generated traces are below the threshold of significance for difference when compared to the real traces. What is missing is a demonstration of how different the traces must be before they are labeled as different. For instance, the authors show that a one hour trace starting at 14:00 is different from a one hour trace starting at 19:00. Is the one hour trace starting at 14:00 different from the one starting at 15:00? Intuitively, the same sort of activities should be happening on a network at those hours of the afternoon. Baring any reasonable explanation of why they should be different, if the statistical measures say that they are different, then we must question the usefulness of those techniques for trace comparison.
- They note that huge data sets do not exactly follow statistical distributions, so they use a random sample of 10000 data points for their comparisons (an accepted method in the statistical community). It seems that the outliers in network traffic are significant, and that ignoring them – especially if they are significant enough to skew a distribution – stands to loose vital characteristics of the network.
- In addition to the two quantitative comparisons, the authors did a qualitative comparison of the Cumulative Distribution Function (CDF); however humans have a tendency to find patterns where they do not exist. This did seem to serve well as a first order comparison to see if two distributions were obviously different.
- As the authors note, only HTTP and FTP were modeled, and a natural extension to their work would be to add additional protocols. We would be concerned, however, that modeling

each protocol by hand does not scale with the growing number of application protocols in use on modern networks, and that failing to model infrequently used protocols, which may be fine when generating traffic for network engineering purposes, will pose problems for the network security community, where an increase of an infrequently used protocol is typically indicative of a new attack. Indeed, the choice to model FTP, with its higher percentage of bytes transferred over DNS, with its higher percentage of packets and sessions, demonstrates this bias in the work as presented.

- Even within the HTTP and FTP models that they used, they chose not to model HTTP clients which use multiple connections per page or the FTP control channel, both of which are bound to have an adverse impact on the use of such models to generate traces for network security testing.
- We also find it odd that their generated HTTP/FTP traffic matched the real traffic when the real traffic was filtered down to just the HTTP and FTP connections. By their own reckoning, a source model is necessary as the individual session traces will behave differently in the presence or absence of other network traffic, so it seems that traces generated in the absence of non-HTTP/FTP traffic should differ from the real traces.

Most recently, Bartoletti and Tang (2005) demonstrated the ability to characterize services based on the differences between vectors to the centroids of clusters representing the services. These representative clusters are produced by clustering thousands of connections to that service into a handful of clusters. Each connection is represented by seven TCP session characteristics. The authors demonstrate that the differences between the centroids of the clusters fall below a standard deviation for clusters within the same service, and are significantly larger between clusters representing different services. Some services are obviously more similar (HTTP and HTTPS) than others (HTTP and FTP). The authors speculate that this method could be used to identify the type of traffic in a tunneled connection, however they do not demonstrate the use of the method for session identification, which is potentially the most interesting application of this approach. Unfortunately, the authors do not have a measure for how many connections are necessary to get a representative cluster. Put another way, they do not know how the method performs given a small number of connections. We would be interested to see how this method compares with our proposed methodology (below), particularly in the area of network (rather than per service) comparison.

Lakhina et al. (2004) conducted the seminal work on applying Principle Component Analysis (PCA) to network traffic. Their work was focused on a single metric with coarse grained measurements. One might imagine that such an approach could be expanded by using

a collection of vectors, each representing a flow, with the dimensions of the vectors provided by the characteristics of the flows. This would, however, present a number of issues, beginning with the representation of numerous multidimensional characteristics in the vector. Secondly, attempting to quantify a similarity from the resulting component vectors would be tenuous. Finally, the existing work utilizing PCA uses coarse grained measurements in large part because of the computational complexity; attempting to run a PCA on such a large number of very large vectors would likely be computationally infeasible.

An alternative approach to comparing network traffic is through the use of anomaly detection systems (such as those used in intrusion detection) to identify how generated traffic differs from real network traffic. It was through the use of anomaly detection systems such as PHAD (Mahoney and Chan 2001) that Mahoney and Chan (2003) identified many of the flaws in the DARPA Intrusion Detection Evaluation dataset. Likewise Hong et al. (2005) were able to use the Q measures from (Javitz and Valdes 1991) for Round Trip Time (RTT) and session duration to validate that the transformed FTP sessions generated by their tool matched real FTP sessions under the same conditions. While anomaly detection is certainly useful in identifying an inconsistency between datasets, it was not designed to provide a quantifiable measure of similarity between the traces. Conceivably, some approaches – such as the Q measures – could be adapted to do so for particular metrics; however, these approaches tend to use very coarse-grained models for the data, which may not provide the fidelity we desire. Other approaches – such as the one PHAD uses – are only useful for finding an anomaly, not measuring its degree.

Beyond looking at just whether generated traffic reproduces real traffic, Hong and Wu (2005) test if it is *effectively* the same, by subjecting both their original and their generated traffic to the Snort open-source, signature-based intrusion detection system (Caswell and Roesch 2004). This is certainly an important, useful, and recommended test in the event one is attempting to closely recreate a particular network environment; however, it is orthogonal to the type of comparison we are performing in this work.

3

Proposed Methodology

Now that we have established the need for quantitative network comparison, we will present a methodology which we believe will serve to provide such a comparison. The approach breaks down into two main stages: Derivation of characteristics and weights, which should need to be done once for any given application of the methodology, followed by the actual comparison of network traces, which can be done for any number of trace pairs. We conclude by enumerating the metrics that we can use to evaluate the correctness of this methodology.

3.1 Methodology summary

Here we provide a quick summary of the steps of our methodology. The subsequent sections will go into greater details of each step and work through an example to provide more context.

3.1.1 Derivation of characteristics and weights

1 Define base cases A subject matter expert defines a number of pairs of network traces which represent the range of network behavior you want to differentiate, and assign similarity values in the range [0..1] to each pair.

2 Define first-order characteristics Define all the characteristics which can be reasonably extracted from the traces for the network protocols under comparison.

Now run each of our base case test pairs through the following steps:

3 Determine distribution function for each characteristic For each defined characteristic, determine how to model its distribution for each trace, and build those models for the first trace.

4 Build the distributions for the second trace Model the same characteristics for the second trace, adjusting characteristics which differ based on the volume of traffic (number of bytes, packets, connections, or similar) by the ratio of traffic between the two traces, such that any given measure of volume is only compared by a single characteristic.

5 Determine similarity function for each characteristic Given the two models for each characteristic in the pair, determine how to compare them to produce a normalized similarity in the range [0..1].

With the above performed on each of the trace pairs for all of the base cases:

6 Find the weights for each characteristic Given all the normalized similarities for all the characteristics of all the base cases, find a weight for each characteristic such that we can produce an overall scaled similarity in the range [0..1] close to the value specified in the first step.

7 Drop non-differentiating characteristics If we are unable to determine a weight for a particular characteristic, we should determine the underlying cause and, if valid, remove the characteristic from our list. For our instantiation on IPv4 traffic, we accomplished this by setting the weight to zero (0).

8 Verify performance using all characteristics Combine the scaled similarities for all the retained characteristics as dictated by the weighting method (for instance, by finding the arithmetic mean). The average scores should be consistent with the intended scores. If it is not, then steps 5 and 6 (and possibly 2 through 4) will need to be revisited.

Now repeat the above steps for higher order characteristics.

9 Determine second-order characteristics and weights Build new characteristics consisting of the first order characteristics that prove useful with respect to any essential (defining) characteristic of the traffic. For IPv4 traffic, this would mean building any characteristic on a per source or dest, IP or port basis, or on the basis of time.

10 Determine higher-order characteristics and weights Continue with this process for higher order characteristics as long as useful characteristics are found.

3.1.2 Comparison of network traces

Once we've defined the characteristics, models, comparison functions, and weights, we can use all of them to find the similarity of any pair of network traces by performing the following steps:

- 1 Build distributions of characteristics** Use the function determined in step 3, above, to build the distributions for all characteristics in each trace. Adjust the distributions of the characteristics in the second trace by any multiplier determined in step 4, above, as necessary.
- 2 Find the scaled similarities for each characteristic** Use the similarity function for each characteristic, determined in step 5, above, to find the normalized similarity between each characteristic. Apply the scaling factor determined in step 6, above, to each normalized similarity.
- 3 Generate overall similarity measure** Combine the scaled similarities appropriately (based on the means in which the weights were derived) to determine the overall similarity measure between the two traces.
- 4 Generate report** The order of the scaled similarities between characteristics, determines which were the most significant contributors for the differences between the two traces. These contributors, plus the overall similarity measure, will be compiled into a user-readable report.

3.2 Detailed Methodology

We now describe the steps of the methodology in detail, along with a simplistic, high-level example, based on comparing how an IPv4 network changes over short time periods.

3.2.1 Derivation of Characteristics and Weights

Define base cases

The first step in the methodology is to collect or build a number of network trace pairs and define the similarity between the each pair of traces. The similarity must be in the range [0..1], and the assignment should be done by a subject matter expert. Each pair will now become a base case.

For our basic example, we will only look at how a network changes over the course of a work day. As such, we'll use two trace pairs from the same network:

- 09:00-10:00 local and 10:00-11:00 local
- 09:00-10:00 local and 03:00-04:00 local

The first pair, we expect to be highly similar, so we'll assign a value of 0.9 to them; the second pair should be fairly different, so we'll assign a value of 0.25 to them.

Define first-order characteristics

The next step is to define the possible first order characteristics that we can derive from the data sources. Specify which of the characteristics are essential in that they are part of what uniquely identifies a connection; these will be characteristics such as host identifiers or service accessed. We will classify the characteristics as total number characteristics (such as total number of bytes sent out of the target network), discrete characteristics (such as the number of packets sent per host), or continuous characteristics (such as the number of connections established in the last 30 seconds).

For our sample instance, we'll just use the following basic IPv4 characteristics:

- Total number of packets (single value metric)
- Total number of connections (single value metric)
- Total number of bytes (single value metric)
- Number of data bytes per protocol/connection dest port (discrete metric)
- Number of packets that arrived at a given minute after the hour (per packet continuous metric)
- Number of SYN flags versus the number of connections over the past w seconds (per packet continuous metric)
- Average duration over the past n connections (per connection initiation continuous metric)
- Control packet rate (per connection close continuous metric)
- Number of packets per IP (continuous metric)
- Number of bytes per IP (continuous metric)

Determine distribution function for each characteristic

For the first trace in each base case test pair, build the distribution of each characteristics defined above:

- The total number characteristics will just be kept as a single value.
- For the characteristics that are labeled as discrete, just keep a table of values.
- We found that just keeping a table of values (either as key, value pairs, or arrays of sorted values) is sufficient for the continuous characteristics as well, at least for our proof of concept on IPv4 traffic. Some prior literature (see Chapter 2) indicates that more complex representations (quadratic functions, fractal distribution, Fourier or Principle Component analysis, wavelet modeling, etc) will be necessary, so they should be considered as analysis and experience dictates.

For our basic example, there is a direct mapping from our characteristics to the functions we'll use for each one. Table 3.1 shows the resulting distributions. These distributions are deliberately convoluted to allow them to remain small enough to use for demonstration purposes, and should not be taken to be representative of any real networks.

Build the distributions for the second trace

For each pair of base case test pairs, build the distributions for the same characteristics of the second trace. If two traces differ only by volume, this will be reflected in the differences between the total packet and byte counts, so we do not want it to also be reflected in the differences between distributions, such as the distribution of packets to privileged services over the past w seconds. To correct for this, use $\frac{\tau_1}{\tau_2}$ as a multiplier for any distributions (or the individual measurements used to build a distribution) built for the second trace¹. This will allow us to compare distributions without worrying about one distribution having twice as much traffic as the other distribution. Characteristics which are invariant with load – for instance error rates which can be represented as a cumulative distribution function (CDF) – will not be scaled. A potential problem with this approach that we will need to remain vigilant of, is that it will also scale traffic that is invariant with load (for instance an automated process that makes a single connection every hour). An alternative approach would be to scale based on the number of active hosts on the target network.

¹where τ_1 and τ_2 are total packet counts (in and out) for distributions based on packet counts, bytes for distributions built on bytes, etc

Characteristic	Pair 1		Pair 2	
	09:00-10:00 local	09:00-10:00 local	09:00-10:00 local	09:00-10:00 local
packets	1508		1508	
connections	17		17	
bytes	458196		458196	
bytes per service	tcp/25	11593	tcp/25	11593
	tcp/80	385629	tcp/80	385629
	tcp/993	18154	tcp/993	18154
	udp/53	42820	udp/53	42820
pkt minute	0	398	0	398
	15	428	15	428
	30	407	30	407
	45	275	45	275
SYN flags conn rate	0	298	0	298
	1	1028	1	1028
	2	159	2	159
	3	23	3	23
ave duration	0	1	0	1
	2	10	2	10
	5	5	5	5
	7	1	7	1
control pkt rate	.25	7	.25	7
	.45	6	.45	6
	.71	1	.71	1
	1	3	1	3
packets per IP	1		1	
	17		17	
	291		291	
	1199		1199	
bytes per IP	129		129	
	5392		5392	
	80935		80935	
	371740		371740	

Table 3.1: Distributions for characteristics of the first trace for each base pair of traces

Characteristic	Pair 1		Pair 2	
	10:00-11:00 local	03:00-04:00 local		
packets	1412		126	
connections	15		2	
bytes	419936		10749	
	tcp/25	9642	tcp/25	10357
bytes per service	tcp/80	354694	tcp/80	0
	tcp/993	17395	tcp/993	0
	udp/53	38205	udp/53	392
	0	412	0	54
pkt minute	16	373	30	72
	30	302		
	44	325		
	0	272	0	12
SYN flags conn rate	1	940	1	114
	2	188		
	3	12		
	0	1	1	1
ave duration	2	10	2	1
	5	3		
	6	1		
	.26	4	0	1
control pkt rate	.48	7	.52	1
	.72	1		
	1	3		
packets per IP	2		1	
	21		15	
	278		110	
	1111			
bytes per IP	10		3	
	10		126	
	10		1899	
	419906		8721	

Table 3.2: Unscaled distributions for characteristics of the second trace for each base pair of traces

Property	Pair 1	Pair 2
packets	1.0680	11.9683
connections	1.1333	8.5
bytes	1.0911	42.6268

Table 3.3: Scaling factors for each pair of traces

Characteristic	Pair 1		Pair 2	
	10:00-11:00 local	03:00-04:00 local		
packets	1412		126	
connections	15		2	
bytes	419936		10749	
bytes per service	tcp/25	10520	tcp/25	441486
	tcp/80	387010	tcp/80	0
	tcp/993	18980	tcp/993	0
	udp/53	41686	udp/53	16710
pkt minute	0	440	0	646
	16	398	30	862
	30	323		
	44	347		
SYN flags conn rate	0	290	0	144
	1	1004	1	1364
	2	201		
	3	13		
ave duration	0	1	1	8
	2	11	2	8
	5	3		
	6	1		
control pkt rate	.26	5	0	8
	.48	8	.52	8
	.72	1		
	1	3		
packets per IP	2		12	
	22		180	
	297		1317	
	1187			
bytes per IP	11		128	
	11		5371	
	11		80948	
	458159		371748	

Table 3.4: Scaled distributions for characteristics of the second trace for each base pair of traces

Table 3.2 shows the unscaled distributions of all the characteristics for the second traces in both of the defined cases. Comparing the counts of bytes, packets, and connections gives us the scaling factors shown in Table 3.3. Given these scaling factors, we can calculate the scaled distributions shown in Table 3.4; for the ease of this example, we rounded the scaled values to the nearest integer value while, in practice, we would retain their full floating point values.

Determine similarity function for each characteristic

For each pair of base case test pairs, calculate the similarity between the two distributions for each characteristic:

- For the total counts, use the formula

$$1 - \frac{|x_1 - x_2|}{x_1 + x_2}$$

as the normalized similarity between the two. This is essentially the Jaccard Coefficient. Alternatively, it is like the percentage similarity ($1 -$ the percentage difference) between the measurements, with the variation that we divide by the sum of the values, rather than the mean of the values, as this will naturally constrain our values in the range $[0, 1]$ rather than $[0, 2]$ (hence, “normalized” similarity), as long as all of our measurements are positive numbers (which in the domain of network measurements, they should be).

- For discrete characteristics, use the average of our normalized similarity between the measurements:

$$1 - \frac{\sum_{i=1}^n \frac{|x_{(i,1)} - x_{(i,2)}|}{x_{(i,1)} + x_{(i,2)}}}{n}$$

where $|x_{(i,1)} - x_{(i,2)}|$ is the absolute value of the difference between the i^{th} measurement of the first and second traces.

- For continuous characteristics, the similarity calculation will depend on the function being used to model the data. We have found two primary ways to compare this data. For the data which is stored as key, value pairs, a weighted value is calculated for each point, taking into consideration the adjacent points. These weighted values can then be compared in the same manner as the discrete characteristics, above. For data which is stored as a sorted array of values, we find the mean of the similarity values for each point on the curve with fewer points against the projected, corresponding point on the other curve. Other functions will require other comparison methods; for instance, if we are using a quadratic function, we should be able to sample the curve at given points, and then use the same similarity formula as we do with the discrete characteristics to find the normalized similarity between

the two curves. More complex representations will require further analysis to determine the proper similarity method.

For our single value metrics we will use the formula

$$1 - \frac{|x_1 - x_2|}{x_1 + x_2}$$

This will be used for the following characteristics:

- Total number of packets
- Total number of connections
- Total number of bytes

For our “number of data bytes per protocol/connection dest port” characteristic we will use the formula for discrete characteristics:

$$1 - \frac{\sum_{i=1}^n \frac{|x_{(i,1)} - x_{(i,2)}|}{x_{(i,1)} + x_{(i,2)}}}{n}$$

For our “number of packets that arrived at a given minute after the hour” characteristic, we will use our approach for ordered continuous characteristics, which involves creating a new of key, value pairs equal to the original set, plus $.5 * value$ to each key adjacent to the current key, and $.25 * value$ for the two keys adjacent to those. The same comparison formula used for discrete characteristics (above) can then be applied.

For our sorted continuous characteristics, we will take the two arrays of distributions in sorted order and compute

$$1 - \frac{\sum_{i=1}^n \frac{|x_{(i,1)} - x_{(j,2)}|}{x_{(i,1)} + x_{(j,2)}}}{n}$$

where n is the number of elements in the smaller array, m is the number of elements in the larger array, $|x_{(i,1)} - x_{(j,2)}|$ is the absolute value of the difference between the i^{th} element of the smaller array and the j^{th} element of the larger array, and $jf = i * \frac{m}{n}$ such that $x_{(j,2)} = x_{(\text{round}(jf),2)}$ if jf is within 0.00001 of that integral value, otherwise, $x_{(j,2)} = \frac{x_{(\text{floor}(jf),2)} + x_{(\text{ceil}(jf),2)}}{2}$. This will work for the following characteristics:

- Number of SYN flags versus the number of connections over the past w seconds
- Average duration over the past n connections
- Control packet rate
- Number of packets per IP

Characteristic	Pair 1	Pair 2
packets	0.9671	0.1542
connections	0.9375	0.2105
bytes	0.9564	0.0458
bytes per service	0.9785	0.1531
pkt minute	0.6154	0.2632
SYN flags conn rate	0.9863	0.8551
ave duration	0.9514	0.6328
control pkt rate	0.9454	0.4238
packets per IP	0.8808	0.6764
bytes per IP	0.2643	0.9985

Table 3.5: Per characteristic similarity values for each of the pairs of traces.

Characteristic	Pair 1 wf	Pair 2 wf	Ave wf	Pair 1 scaled sim	Pair2 scaled sim
packets	0.9306	1.6210	1.2758	1.2339	0.1968
connections	0.9600	1.1875	1.0737	1.0066	0.2261
bytes	0.9410	5.4534	3.1972	3.0579	0.1466
bytes per service	0.9198	1.6325	1.2761	1.2487	0.1954
pkt minute	1.4624	0.9497	1.2061	0.7422	0.3175
SYN flags conn rate	0.9125	0.2924	0.6024	0.5942	0.5151
ave duration	0.9460	0.3951	0.6706	0.6379	0.4243
control pkt rate	0.9520	0.5898	0.7709	0.7288	0.3268
packets per IP	1.0218	0.3696	0.6957	0.6128	0.4706
bytes per IP	3.4047	0.2504	1.8275	0.4831	1.8248

Table 3.6: Per characteristic weighting factors (wf) for each of the pairs of traces to bring the corresponding similarity to 0.9 for pair 1 and 0.25 for pair 2, followed by the average weighting factor, and then the scaled similarities for each characteristic in each pair of traces.

- Number of bytes per IP

Given these formulas, we take the data in Table 3.1 and Table 3.4 and find the similarity values for each characteristic, using the code in appendix A, producing the per characteristic similarities shown in Table 3.5.

Find the weights for each characteristic

Once we have performed the above steps for each of the trace pairs for the base cases, we must consider the usefulness of each characteristic, given its behavior across all the base cases. In the process of doing so, we should find adjustment factors for each characteristic to determine how much influence that characteristic has on the overall similarity value. The normalized similarity will be known as the scaled similarity once multiplied by the adjustment factor. Any process to determine – and later apply – these adjustment factors is acceptable, as long as it is consistent.

For our basic example, we find the multipliers to bring each similarity value to the goal values of 0.9 for pair 1 and 0.25 for pair 2. These multipliers are shown in the first two columns of Table 3.6. We then take the average of these two values to find the per characteristic weight, which is the next column in the table. These weights are multiplied against the normalized similarity values shown in Table 3.5 to produce the scaled similarities, which are the right two columns in Table 3.6.

Alternative methods may be used to find and apply these weights, and the proper method will need to be determined by a subject matter expert. For our sample implementation on IPv4 traffic, we found that linear regression works best. It is conceivable that the method may even be non-linear in nature; however, our experiments with non-linear scaling indicated that it overfit the weights.

Drop non-differentiating characteristics

If we are unable to find an adjustment factor which allows the normalized similarity to reach the target similarity value, for a significant number of base cases², we must determine if that is

1. due to a failure of the distribution algorithm used (for instance if it tells us that two traces are similar when they are not; however, equally likely in other situations),
2. a failure for a characteristic to change between different traces (be non-differentiating), or
3. a data failure in two traces being more or less similar in some respect than intended.

If a characteristic does not differentiate a significant number of base cases (either because it is invariant, or it is random), then drop it.

If we look at our basic example, we see the overall similarities are fairly close to our goals for both pairs. Yet if we look at the per characteristic similarities, we see that the similarities for the number of bytes per IP are far askew from the goals for each pair; in fact, the similarity value for the first pair is close to the goal for the second pair and the similarity of the second pair is close to the goal for the first. In this case, a quick look at the underlying data shows that the characteristic itself is not indicative of the overall similarity of the traces, so we drop it from the list of characteristics we use.

²Currently, determining how many makes it significant is an open problem.

Verify performance using all characteristics

Take the scaled similarity for all the retained characteristics and combine them as dictated for the weighting method being used. The average scores should be consistent with the intended scores. If it is not, then steps 5 and 6 (and possibly 2 through 4) will need to be revisited.

For our basic example, combining the per characteristic scaled similarities is just done by finding the weighted mean of the scaled similarities: divide the sum of the scaled similarities by the sum of the weights. By doing this, we get 0.9159 for the first pair, and 0.2618 for the second: both very close to our goals of 0.9 and 0.25, respectively.

Other weighting methods, such as the linear regression method used in our main implementation, or a non-linear weighting, may require different ways of combining the individual scaled similarities.

Determine second-order characteristics and weights

For any of the first order characteristics that were useful, repeat steps 2 through 7 above, treating it as a second order characteristic to each member of the set of essential characteristics (which was specified in the list of characteristics). If the second order characteristic proves useful, integrate it with the others as above.

Extending our example, we would consider each of our ten characteristics with respect to each essential characteristic of an IPv4 connection. For example, we would consider the following second-order characteristics based on the “Number of packets” metric:

- Number of packets per second after the minute
- Number of packets per minute after the hour
- Number of packets per local hour
- Number of packets per UTC hour
- Number of packets per source IP
- Number of packets per destination IP
- Number of packets per source port
- Number of packets per destination port

Characteristic	Pair 1			Pair 2		
	09:00-10:00 local			09:00-10:00 local		
duration per sec	0	2	2	0	2	2
		5	1		5	1
	15	2	3	15	2	3
		5	2		5	2
	30	0	1	30	0	1
		2	1		2	1
		5	1		5	1
		7	1		7	1
	45	2	4	45	2	4
		5	1		5	1
duration per dest port	tcp/25	2	1	tcp/25	2	1
		5	1		5	1
	tcp/80	0	1	tcp/80	0	1
		2	3		2	3
		5	3		5	3
		7	1		7	1
	tcp/993	2	2	tcp/993	2	2
		5	1		5	1
	udp/53	2	4	udp/53	2	4

Table 3.7: Distributions on the first traces of each pair for the second-order characteristics

Following the same approach for the other nine basic characteristics will give us 80 characteristics to consider. There is some redundancy to eliminate though; for instance, “Number of packets per IP per source IP” doesn’t make any sense. We come across these artifacts by virtue of modeling the essential characteristics with our basic characteristics.

We choose to consider only higher order characteristics of lower order ones which proved useful, and limit them to be higher order to the essential characteristics only as a way of limiting our search space in this combinatorial problem. Granted, there is the possibility that there might be a higher-order characteristic this does not find, for instance between the relationship between the number of packets in the past n connections and the distribution of duration times in a given trace; however, if we did that, we would be more likely to find coincidences that have no basis in reality, or obvious relationships that provide no additional insight.

Rather than expand our example by showing the almost 80 second-order characteristics, we’ll just look at two of them: “Average duration over the past n connections per second after the minute” and “Average duration over the past n connections per destination port”. For the distribution functions, we’ll just extend our associated list structure by adding the new key (second after the minute or destination port) and breaking out the associated values (which is itself an associated list). By doing this on the first trace, we get the distributions shown in Table 3.7. We then build the distributions for the second trace in each pair, getting the unscaled distributions shown in Table 3.8 then, by applying the same scaling factors from Table 3.3 we

Characteristic	Pair 1			Pair 2		
	10:00-11:00 local			03:00-04:00 local		
duration per sec	0	2	2	15	1	1
		5	1			
	15	2	3	30	2	1
		5	1			
	30	0	1			
		2	1			
		6	1			
	45	2	4			
		5	1			
duration per dest port	tcp/25	2	1	tcp/25	2	1
		5	1			
	tcp/80	0	1	udp/53	1	1
		2	3			
		5	2			
		6	1			
	tcp/993	2	2			
	udp/53	2	4			

Table 3.8: Unscaled distributions on the second traces of each pair for the second-order characteristics

Characteristic	Pair 1			Pair 2		
	10:00-11:00 local			03:00-04:00 local		
duration per sec	0	2	2	15	1	8
		5	1			
	15	2	3	30	2	8
		5	1			
	30	0	1			
		2	1			
		6	1			
	45	2	5			
		5	1			
duration per dest port	tcp/25	2	1	tcp/25	2	8
		5	1			
	tcp/80	0	1	udp/53	1	8
		2	3			
		5	2			
		6	1			
	tcp/993	2	2			
	udp/53	2	5			

Table 3.9: Scaled distributions on the second traces of each pair for the second-order characteristics

Characteristic	Pair 1	Pair 2
duration per sec	0.9538	0.1667
duration per dest port	0.5546	0.2

Table 3.10: Normalized similarities of second-order characteristics

Characteristic	Pair 1 wf	Pair 2 wf	Ave wf	Pair 1 scaled sim	Pair 2 scaled sim
duration per second	0.9436	1.5000	1.2218	1.1654	0.2036
duration per dest port	1.6230	1.2500	1.4365	0.7966	0.2873

Table 3.11: Per characteristic weighting factors (wf) for each of the pairs of traces to bring the corresponding similarity to 0.9 for pair 1 and 0.25 for pair 2, followed by the average weighting factor, and then the scaled similarities for each second-order characteristic in each pair of traces.

get the scaled distributions shown in Table 3.9.

To do the actual comparison, we'll find the mathematical mean of the similarities of the sorted arrays which have the same outer key (seconds after the minute or destination port). The code to do this calculation is shown in appendix A. Using this approach, we get the normalized similarities shown in Table 3.10. Following the same process as we do for the first order characteristics, we get the per-characteristic weighting factors and average weighting factors shown in Table 3.11. Applying these back to the normalized similarities shown in Table 3.10 gives us the scaled similarities shown in Table 3.11. Both characteristics appear to work as intended, so we retain them. When we combine these scaled similarities with the others from Table 3.6, we get overall similarities of 0.8807 for the first pair of traces and 0.2465 for the second pair.

Determine higher-order characteristics and weights

Continue with this process for higher order characteristics as long as useful characteristics are found. For our basic example, we might consider the “Average duration over the past n connections per second after the minute per source port” and “Average duration over the past n connections per destination port per destination IP”. The implementation of these characteristics follows directly from the above, so we will not go into it here.

3.2.2 Comparison of IP Network Traces

Once we perform all the steps in the previous section, we know the characteristics, equations, and weights to use when comparing two network traces. Now we are ready to apply them to any number of trace pairs, using the following steps:

Build distributions of characteristics

Use the function determined in step 3, above, to build the distributions for all characteristics in each trace. Adjust the distributions of the characteristics in the second trace by any multiplier determined in step 4, above, as necessary.

Table 3.12 shows the raw distributions for two pairs of traces we will use to illustrate the application of the methodology to determine the similarities between the pairs. We adjusted the figures for the second trace by the ratio of connections, packets, or bytes between the first and second traces (not shown for brevity).

Find the scaled similarities for each characteristic

Use the similarity function for each characteristic, determined in step 5, above, to find the normalized similarity between each characteristic. Apply the scaling factor determined in step 6, above, to each normalized similarity.

Extending our example we used the scaling factors shown in Table 3.6 and Table 3.11 to find the scaled similarities shown in Table 3.13.

Generate overall similarity measure

Combine the scaled similarities appropriately (based on the means in which the weights were derived) to determine the overall similarity measure between the two traces.

For our example, we find the weighted mean similarity value for the first pair to be 0.8881 and 0.5342 for the second pair. This indicates that the first pair of our test was about as similar in network activity as the first basemode we built our weights from. The traces of the second pair for this test were much less similar overall, to those pairs, but more similar to each other than the second basemode that the weights were built from.

Generate report

The order of the scaled similarities between characteristics, determines which were the most significant contributors for the differences between the two traces. These contributors, plus the overall similarity measure, will be compiled into a user-readable report.

We will not actually compile a complete report for the purposes of this example; the interested reader can see such reports from our sample implementation in Appendix M.1. Table 3.13 gives a good idea of the type of data that would appear in such a report. Given that Table 3.13 is small enough, it is easy enough to see that trace pair 1 was most similar due to bytes transmitted, and least similar in the number of SYN flags versus the number of connections

Characteristic	Pair 1				Pair 2				
	08:30-09:30 local		09:30-10:30 local		08:30-09:30 local		22:30-23:30 local		
packets	1903		2021		1903		666		
connections	20		22		20		6		
bytes	619384		682845		619384		124742		
bytes per service	tcp/25	104835	tcp/25	92913	tcp/25	104835	tcp/25	93825	
	tcp/80	298572	tcp/80	364504	tcp/80	298572	tcp/80	16715	
	tcp/443	194035	tcp/443	201836	tcp/443	194035	tcp/443	1257	
	udp/53	21942	udp/53	23592	udp/53	21942	udp/53	12945	
pkt minute	0	492	0	512	0	492	0	182	
	15	462	15	538	15	462	15	201	
	30	487	30	478	30	487	30	174	
	45	462	45	493	45	462	45	109	
SYN flags conn rate	0	194	0	195	0	194	0	57	
	1	1007	1	1035	1	1007	1	589	
	2	477	2	502	2	477	2	20	
	3	225	3	289	3	225			
ave duration	0	10	0	11	0	10	0	2	
	1	6	1	8	1	6	1	3	
	3	2	3	1	3	2	5	1	
	6	2	7	2	6	2			
control pkt rate	.25	4	.26	5	.25	4	.33	2	
	.49	7	.51	8	.49	7	.67	3	
	.76	5	.74	4	.76	5	1	1	
	1	4	.99	5	1	4			
packets per IP	192		201		192		123		
	496		373		496		222		
	523		440		523		321		
	692		1007		692				
duration per sec	0	0	2	0	0	3	0	0	2
		1	2		1	1		1	2
		3	1		0	3		3	1
	15	0	3	15	1	3	15	0	3
		1	1		3	1		1	1
		6	1		7	1		6	1
	30	0	2	30	0	3	30	0	2
		1	1		1	3		1	1
		3	1		0	2		3	1
	45	6	1	45	1	1	45	6	1
		0	3		7	1		0	3
		1	2					1	2
duration per dest port	tcp/25	0	3	tcp/25	0	2	tcp/25	0	3
		1	1		1	2		1	1
		6	1		0	3		6	1
	tcp/80	0	2	tcp/80	1	2	tcp/80	0	2
		1	1		7	1		1	1
		3	1		1	1		3	1
	tcp/443	1	1	tcp/443	3	1	tcp/443	1	1
		3	1		7	1		3	1
		6	1		0	6		6	1
	udp/53	0	5	udp/53	1	3	udp/53	0	5
		1	3					1	3

Table 3.12: Distributions for all characteristics from all traces

Characteristic	Pair 1	Pair 2
packets	1.2374	0.6615
connections	1.0226	0.4956
bytes	3.0414	1.0719
bytes per service	1.2119	0.4383
pkt minute	1.1711	1.1111
SYN flags conn rate	0.5915	0.5121
ave duration	0.6487	0.5388
control pkt rate	0.7501	0.6834
packets per IP	0.6172	0.5701
duration per sec	0.8820	0.7363
duration per dest port	0.7506	0.3541

Table 3.13: Scaled similarities of all characteristics

over the past w seconds. Likewise, trace pair 2 was most similar in the number of packets that arrived at a given minute after the hour, and least similar in the average duration over the past n connections per destination port.

3.3 Success metrics

To test the hypothesis that the above methodology provides us a quantitative similarity between two network traces, we have identified six metrics to test for success:

1. A network trace and an anonymized version of the same trace show no difference.
2. The same network at adjacent times (without major sociological difference, such as the start of a workday) shows some minimal difference.
3. The same network at the same time, different weeks, shows some minimal difference.
4. The same network at disparate times (on/off hours) shows more significant difference.
5. Similar networks show some difference at the same time, and different networks (edu/edu vs edu/com) show larger difference.
6. Completely different traces (disjoint networks, disjoint times, disjoint services) show minimal similarity.

If an instance of the proposed methodology is able to pass all six success metrics, we can be confident that the methodology is sound, as it means that the methodology is producing similarity values which are relatively consistent with what a subject matter expert would assign.

4

Initial steps

Before performing any actual comparisons, we have some initial steps to the methodology we need to follow. For starters, we need to enumerate the characteristics we will use as the basis for our comparisons. With that we need to define the base cases we'll use both to calibrate our weights and evaluate the success of our methodology. Then we can build the data that we will use for initial testing and calibration. This chapter covers these initial steps.

4.1 Characteristics for comparing IPv4 networks

The first step in the methodology is to define what characteristics we might be able to extract from the network traces to compare the data. For our example instantiation, we are looking at comparing IPv4 networks, as they are the most widely deployed networks today, and hence the easiest source of data. We are also using the TCP, UDP, and ICMP connection layer data, which is available on almost all of the collected packets. We are not extracting application level characteristics for protocols such as SMTP, Telnet, HTTP, POP, IMAP, etc, although it is possible (if time consuming) to do so. We are also not looking at the distribution of content in the data portions of the traffic, although – again – it is possible to do so.

The enumeration of characteristics is shown in Appendix B. While it is based on our research in the field of network intrusion detection (Brugger 2007a), we feel that it should be generally applicable.

4.2 Base cases

At the end of chapter 3 we presented six metrics to test for success of our methodology. These metrics are rather broad, so we need some specific goals to test for success and calibrate the appropriate characteristics and weights. We have identified 17 specific base cases that we can use to this end. Of these, five can be generated from a single trace, making them useful for initial testing and calibration. Using six others, we can cover all six of the success metrics.

The base cases are as follows:

1. One hour of data and anonymized version of that hour (should score 1.0)
2. One hour of data and the same hour repeated with the timestamps modified to appear to be one hour later (should score 1.0)
3. One hour of data and the same hour repeated one hour later with a few connections added, a few removed, most shifted slightly in time, some host pairs changed (should score approx 0.9)
4. The same pair with even more modifications (should score lower)
5. One hour of data from one network, and the same hour of data generated (by manually manipulating real data as necessary) to share no similarity (size, protocol mix, topology, etc) with the first network trace (score should approach 0.0 – it will not be zero unless one network has traffic and the other does not)
6. Two adjacent hours of traffic from the same network in mid-day (score should be > 0.75)
7. Two adjacent hours of traffic from the same network in off-hours (score should be > 0.75)
8. The same mid-day hour of traffic from the same network on adjacent weeks (score should be > 0.75)
9. The same off-hour of traffic from the same network on adjacent weeks (score should be > 0.75)
10. One mid-day hour and one off-hour of traffic from the same network (score should be lower than the two above scores)
11. One hour of data from one network, and the same hour of data from a different network that is similar in size, topology, policy, user base, etc, such that we should expect the traffic to be similar in composition (score should be > 0.5 and < 0.75)

12. One hour of data from one network, and the same hour of data from a different network that differs in a single significant way (size, topology, policy, user base, etc) from the first such that we should expect the traffic to share some similarities, but still be less similar in composition to the above (score should be > 0.25 and $<$ above)
13. One hour of data from one network, and the same hour of data from a significantly different network (score should be < 0.25)
14. One day of data from a network, and one day of data from the same network on the same day of the following week (score should be > 0.75)
15. One day of data from one network, and the same day of data from a similar network (score should be > 0.5 and < 0.75)
16. One day of data from one network, and the same day of data from a slightly different network (score should be > 0.25 and < 0.5)
17. One day of data from one network, and the same day of data from a very different network (score should be < 0.25)

Of these, we use the following five for initial testing and calibration:

1. One hour of data and anonymized version of that hour (should score 1.0)
2. One hour of data and the same hour repeated with the timestamps modified to appear to be one hour later (should score 1.0)
3. One hour of data and the same hour repeated one hour later with a few connections added, a few removed, most shifted slightly in time, some host pairs changed (should score approx 0.9)
4. The same pair with even more modifications (should score lower)
5. One hour of data from one network, and the same hour of data generated (by manually manipulating real data as necessary) to share no similarity (size, protocol mix, topology, etc) with the first network trace (score should approach 0.0 – it will not be zero unless one network has traffic and the other does not)

The first of these base cases will ensure that first success metric is met.

We use six additional base cases to ensure that we can test for all of our success metrics:

1. Two adjacent hours of traffic from the same network in mid-day (score should be > 0.75)

2. The same mid-day hour of traffic from the same network on adjacent weeks (score should be > 0.75)
3. One mid-day hour and one off-hour of traffic from the same network (score should be > 0.5 and < 0.75)
4. One hour of data from one network, and the same hour of data from a similar network (score should be > 0.5 and < 0.75)
5. One hour of data from one network, and the same hour of data from a slightly different network (score should be > 0.25 and < 0.5)
6. One hour of data from one network, and the same hour of data from a very different network (score should be < 0.25)

Each of the above base cases correspond to one of the success metrics, with the exception of the fourth and fifth base cases, which test the two assertions in success metric five. We would have liked to test all seventeen base cases, however this was seen as too large an effort for the scope of this dissertation, and leaves us room for future work in this area.

4.3 Initial test data

The first five base cases can be generated using only a single network trace, making them a great starting point for building the tools to instantiate and test the proposed methodology. This is particularly true as collecting the data from different types of networks will take some time.

Appendix C describes in great detail exactly how we created the pcap network trace pairs for these five base cases. All the tools attributed to the author are available at (Brugger 2007c).

5

Single Value and Discrete Metrics on Initial Base Cases

Given the first five base cases, we proceeded to develop scripts to extract the characteristics, find the similarity score for each characteristic, and develop the weights for each similarity to determine the overall similarity score. The first five base cases, as covered in the previous chapter, are all constructed by hand based on one hour of real network data. By using constructed data, we know what to expect, reducing surprises during the development of the methodology.

This chapter will begin by reviewing our problems with count and discrete metrics, turn to the actual metrics that we will be looking at, and then cover the normalized similarity, and wrap up with our scaled similarity calculations.

All numbers in this chapter have been rounded to four digits after the decimal place for legibility. Actual calculations were performed using single-precision floating point numbers.

5.1 The trouble with count and discrete metrics

We made a lot of wrong turns getting to the point you see presented in this chapter. The basic things we have learned are:

- Some metrics may be misleading, and it will be easier to calculate a scaled similarity without them.
- Linear scaling works better than non-linear (quadratic) scaling as the non-linear approach tends to over-fit, and the metrics that seemed misleading before can once again be useful.

- Per IP metrics can be handled if the observed network is anonymized by mapping the hosts of that network to the hosts of the observed network in the trace it is being compared with.
- By not keying off of specific IP addresses and instead treating Per IP metrics as continuous distributions, we can handle traces that are completely anonymized, as well as networks that are much larger.
- Besides the six count metrics, there are four ratio metrics that we should evaluate as single value metrics.

We have included this chapter, as it originally appeared, as Appendix D. That chapter covers the first three points, above, in depth. In short, when we took the normalized similarities from our counts and calculated a second-order quadratic equation to try to fit them to the expected similarity, the “bytes in” and “bytes out” metrics values were so extreme, the quadratic ended up being an inverted parabola. We found that dropping them allowed the other four count metrics to more easily scale to the intended result (as defined by our base cases). When we then looked at discrete metrics, we had many more inverted parabolas, so we replaced the quadratic scaling equations with linear scaling equations (and reintroduced “bytes in” and “bytes out”), and found that our scaled metrics closed mapped to our expected metrics.

We also originally looked at six “Per IP” metrics as discrete metrics:

- Source IP based on the number of packets
- Source IP based on the number of connections
- Source IP based on the number of bytes transferred
- Destination IP based on the number of packets
- Destination IP based on the number of connections
- Destination IP based on the number of bytes transferred

We found that this caused problems in the first basecase, which compares a trace to an anonymized copy of the same trace. Since one of the goals of the methodology is to be agnostic to the actual addresses used on the network, these two traces are supposed to be perfectly similar, but the per IP metrics were causing them to be different. To address this, we developed a script that compared every active host on the observed network of the first trace to every active host on the observed network of the second trace, and mapped the most similar hosts to one another. This

addressed the problem in the short-term, but with this $O(n^2)$ operation¹, it seriously limited the size of networks we could examine.

A bigger problem arose when we went to collect data to evaluate the methodology on real-world network data: most sites insisted on anonymizing both the internal and external addresses in any traces. The $O(n^2)$ operation was painful enough – we did not think a $O(n^2 + m^2)$ operation² was feasible. To address this, we decided not to key off the specific IP addresses, and instead to treat these six metrics as continuous characteristics, where we can just compare the metrics as curves of sorted values. These metrics are now covered in more detail in chapter 7.

Finally, in looking at connection characteristics, we found four that we originally treated as continuous characteristics, until we realized that they could only have a value of 0 or 1, which lent them to being treated as ratios, which then allowed us to use the same techniques as we did for the count metrics.

5.2 Single value and discrete metrics

We have 17 single value and discrete metrics:

1. Bytes sent into the target network
2. Bytes sent out from the target network
3. Packets sent into the target network
4. Packets sent out from the target network
5. Connections into the target network
6. Connections out from the target network
7. SYN-ONLY connection rate
8. SYN-ACK connection rate
9. Idle connection rate
10. Half-open connection rate
11. Protocol/dest port (or ICMP type/code) based on the number of packets

¹Where n is the maximum number of hosts on the observed networks

²Where m is the maximum number of hosts external to the observed network that had some communication with hosts on the observed network

12. Protocol/dest port (or ICMP type/code) based on the number of connections
13. Protocol/dest port (or ICMP type/code) based on the number of bytes transferred
14. TCP/UDP Source Port based on the number of packets
15. TCP/UDP Source Port based on the number of connections
16. TCP/UDP Source Port based on the number of bytes transferred
17. TTL Values

This section will briefly look at exactly what each metric entails and why we are including it.

5.2.1 Count metrics

The count metrics are fairly straightforward: they represent the total count of the given metric for the entire trace with respect to the network being observed. We only consider Internet traffic; we did not examine traffic where the source and destination were both internal or external to the target network.

Bytes in

The bytes sent into the target network (or “Bytes in” for short) is the total number of bytes (including headers) sent into the observed network for the duration of the observation (trace). Headers for the logical layer (Ethernet, frame relay, etc) are not included.

Bytes out

The bytes sent out from the target network (or “Bytes out” for short) is the total number of bytes (including headers) sent out from the observed network for the duration of the observation (trace). Headers for the logical layer (Ethernet, frame relay, etc) are not included.

Packets in

The packets sent into the target network (or “Packets in” for short) is the total number of packets sent into the observed network for the duration of the observation (trace).

Packets out

The packets sent out from the target network (or “Packets out” for short) is the total number of packets sent out from the observed network for the duration of the observation (trace).

Connections in

The connections into the target network (or “Connections in” for short) is the total number of connections into the observed network for the duration of the observation (trace). Chapter 6 covers exactly what constitutes a connection in detail. In short, it is any logical group of packets between a pair of hosts. We only count TCP connections that we see initiated, so we do not include TCP connections that were already in progress when observation began. For this chapter, we count every UDP packet as a connection. The connection count includes unsuccessful connection attempts (connections that do not make it to the ESTABLISHED state).

Connections out

The connections out from the target network (or “Connections out” for short) is the total number of connections out from the observed network for the duration of the observation (trace). Chapter 6 covers exactly what constitutes a connection in detail. The description of Connections in, above, covers a brief description that should suffice for this chapter.

5.2.2 Ratio metrics

As described in chapter 7, we have a number of connection metrics that are output at the conclusion of each connection, which we term “Per connection close” metrics. Four of those specify whether or not the connection is in one of four anomalous states at the time it closes: SYN-ONLY, SYN-ACK, Idle, or Half-open. Since all of the other “Per connection close” metrics are continuous, we initially treated these in the same manner. As we show in the discussion of SYN-ONLY connections, we eventually realized that, as these were binary in value (0 or 1), they would be better handled as ratios. This way we can see what portion of the closed connections we observe ended in one of those states. By doing this, we can use the same comparison formulas as we do for the count metrics, above. This section briefly reviews the four ratio metrics.

SYN-ONLY connections

A connection is considered to be in the SYN-ONLY state when a SYN was sent by the initiator, but never acknowledged. This frequently occurs when someone tries to connect to a service that is not offered by the remote machine, or is blocked by a firewall. Such a connection may be benign (trying to connect to a server that moved or a typo) or malicious (probing or worm activity).

As noted above, this metric is output at the end of each connection, so we initially treated them as continuous characteristics. This mind set was reinforced after the analysis of

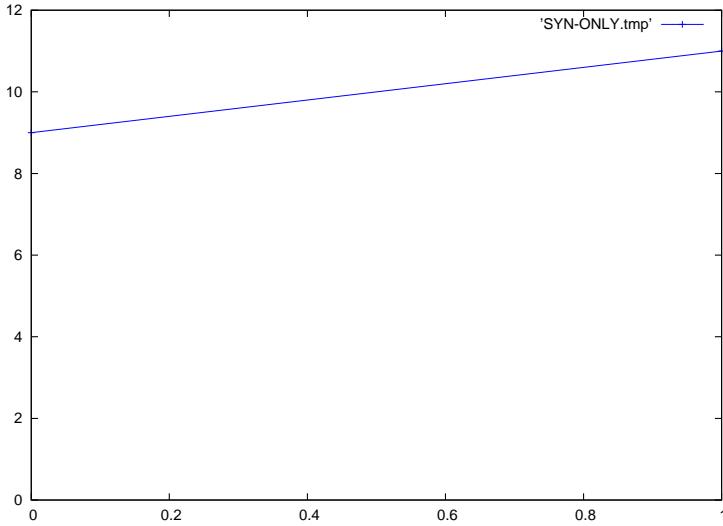


Figure 5.1: Plot of SYN-ONLY values as continuous characteristic for one hour of sample data.

“Connection errors” in section H.1.6 showed that, despite the small number of values observed in the sample data, Figure H.75 showed a distinct continuous line which could be modeled and compared. Given this, when presented with the plot of SYN-ONLY values in Figure 5.1, the idea that this was a continuous function did not seem far fetched.

It was only when we realized that “Connection errors” and the other “Per connection close” metrics may not have many values in the sample data, but in theory did not have an upper bound to the range of their independent variable, whereas SYN-ONLY could only be a 0 or a 1. A metric with only two possible values is not very continuous. It is not even very useful to think of it as a discrete metric. Really, probably the easiest way to think of this is as a single value metric – in this case, with 11 connections that ended in the SYN-ONLY state out of 20 total connections, we have $11/20 = 0.55$.

SYN-ACK connections

A connection is considered to be in the SYN-ACK state when the remote server ACKed the initial SYN, however the SYN-ACK was never ACKed to establish a connection. Common explanations for this would be sudden loss of connectivity, or probing activity (once the presence of the server has been established, there is no need to continue communicating with it).

By the same logic as we have for SYN-ONLY, we look at this as a ratio of the number of connections that ended in this state over the total number of connections.

Idle connections

A connection is considered to be in the Idle state when no activity has been seen on this connection in a time interval exceeding some threshold. The longer a connection sits idle, the more likely that one or both sides have silently closed the connection (say, by going into sleep or standby).

By the same logic as we have for SYN-ONLY, we look at this as a ratio of the number of connections that ended in this state over the total number of connections.

Half-open connections

A connection is considered to be in the Half-open state when one side of the connection initiated a close with a FIN, but the connection did not completely close (one or more of the FIN from the other side or the ACKs to those FINs was not observed).

By the same logic as we have for SYN-ONLY, we look at this as a ratio of the number of connections that ended in this state over the total number of connections.

5.2.3 Discrete metrics

Discrete metrics are those that we have measurements for at specific points, and will compare to each other at those points (inserting zero for a measurement not seen in a trace, which was seen in the other), averaging the results to generate a single normalized similarity.

We have seven such metrics, which look at the TTL values, and at the source and dest ports from the vantage of number of packets, connections, and bytes transferred. One might argue that these metrics are second-order metrics, since they take other first-order metrics at look at them on a “per” something basis. We will consider them first-order metrics, however, we can think of no other way to examine the source ports, and it is important to us to have them included as first-order metrics such that we can evaluate them for inclusion as higher-order metrics.

Packets per protocol/dest port

This metric looks at the number of packets sent to a given protocol/dest port. This allows us to look at the number of packets to the same port on TCP or UDP separately. Since ICMP does not have dest ports, we use the combination of the ICMP type and code. We do not generate counts for packets that are not TCP, UDP, or ICMP.

This metric gives us a very rough approximation of how heavily used each protocol is. It is a rough approximation, because the determination of dest port is made on a per packet

basis, so any responses from the service will use the client's source port as the dest port for that packets, and due to ACK packets, the number of packets going in each direction will be fairly equivalent. We rely on the aggregation from numerous connections to create spikes on the ports providing services.

Connections per protocol/dest port

This metric looks at the number of connections created to a given protocol/dest port. This allows us to look at the number of connections to the same port on TCP or UDP separately. Since ICMP does not have dest ports, we use the combination of the ICMP type and code. We do not generate counts for connections that are not TCP, UDP, or ICMP.

This will be a good measure of relative use of the different services to or from the observed network, as it does not have the problem that the packet count does in counting the reply packets (where the original source port becomes the destination port), however it does not give us an idea of what the relative amount of traffic was to those services. For example, a site may send and receive a lot more mail than file transfers, however file transfers may make up the bulk of the traffic (both on a packet count and a byte count basis) because they are much larger.

Bytes per protocol/dest port

This metric looks at the number of bytes transferred (including headers) to a given protocol/dest port. This allows us to look at the number of bytes to the same port on TCP or UDP separately. Since ICMP does not have dest ports, we use the combination of the ICMP type and code. We do not generate counts for packets that are not TCP, UDP, or ICMP.

This metric has the same disadvantage as the packet count per dest port in that replies from the server will use the client's port as the dest port, however it is compounded here in that the number of bytes to and from the server may be very asynchronous: requests to the server may be succinct, and the replies large in the number of bytes. Given such a mismatch, we can not expect that the cumulative traffic to server ports will separate them from client ports. More likely, such a separation will be made simply on the basis that servers tend to use privileged ports (those below 1024), and clients tend to use non-privileged ports (1024 and up). Even without this separation, it should still be a useful metric for comparing networks that are not radically different, particularly looking at how a network changes over time.

Packets per protocol/source port

This metric looks at the number of packets received from a given protocol/source port. This allows us to look at the number of packets to the same port on TCP or UDP separately.

We only generate counts for packets that are TCP or UDP.

Due to the symmetry in packet counts between the client and the server, discussed above in the section about packet counts to dest ports, we suspect that the packet counts on a per source basis will look extremely similar. This gives rise to the idea that we may be measuring the wrong thing here: right now we are calculating our per source and dest port counts based on the values seen in the packet. It is likely much more informative if we were to look at these based on the source and dest at connection initialization time. For now, we will see how well what we have works and leave this possibility for future work.

Connections per protocol/source port

This metric looks at the number of connections received from a given protocol/source port. This allows us to look at the number of connections from the same port on TCP or UDP separately. We only generate counts for connections that are TCP or UDP.

As discussed above, this will give us an actual look at the ports used by clients in connecting to servers. It should be informative in that a select number of services use the same source port for all connections. Additionally, services that just let the OS pick the source port will tend to use sequentially increasing port numbers, starting at 1024 and going up to some max before wrapping around. To this end, the source port is indicative of how many connections that host makes. Because of this property, we also look at the source ports as continuous characteristics in chapter 7.

Bytes per protocol/source port

This metric looks at the number of bytes transferred (including headers) received from a given protocol/source port. This allows us to look at the number of bytes from the same port on TCP or UDP separately. We only generate counts for packets that are TCP or UDP.

As discussed above, this metric has the same problems as bytes per dest port, compounded due to the asynchronicity, so the same caveats apply.

TTL Values

We look at the Time To Live (TTL) on all of the packets that pass by the observer. The TTL that is observed is determined by a combination of the packet sender's operating system and the logical distance that the packet has traveled before it is observed. While the first factor – the mix of origin operating systems – will produce a consistent plot, there is a good deal of noise in the second factor due to the dynamic nature of packet routing. For this reason we anticipate

that this metric will be better modeled as a continuous characteristic. We do this in chapter 7, which includes a more detailed discussion of what TTL plots look like.

5.3 Normalized similarity

Once we have the distributions for all of our metrics on all of the traces of interest, we pass the two distributions to the `normCompare.pl` script, written by the author. This script takes the union of characteristics between the two distributions and outputs a normalized similarity for each one.

For single value characteristics we use the formula

$$1 - \frac{|x_1 - x_2|}{x_1 + x_2}$$

with the additional caveat added for ratio metrics that if $(x_1 + x_2) = 0$, then the normalized similarity is defined to be one (1). Otherwise, the value is limited to the range $[0, 1]$ (hence, “Normalized similarity”).

For discrete characteristics, we first adjust the packet, connection, or byte count by $\frac{\tau_1}{\tau_2}$, where τ_1 is the count for the first trace, and τ_2 is the count for the second trace. This allows for the scale of packets, connections, or byte counts to be the same for both discrete distributions; that way we are actually comparing the distributions, not the scales of the networks, which is done with the packet, connection, and byte counts themselves.

Once that is done, we compute

$$1 - \frac{\sum_{i=1}^n \frac{|x_{(i,1)} - x_{(i,2)}|}{x_{(i,1)} + x_{(i,2)}}}{n}$$

where $|x_{(i,1)} - x_{(i,2)}|$ is the absolute value of the difference between the i^{th} measurement of the first and second traces.

The results of all the normalized similarity computations are shown in Table 5.1. As you can see, some of the normalized similarity values vary significantly between the baselines, and some are invariant. Some even increase in value on trace pairs that are supposed to be less similar than previous ones. We will look at these cases in more detail in the next section.

5.4 Scaled similarity

If we look at the mean of the normalized similarities in Table 5.1, we see that they are close to, but not quite the expected values of $\{1.0, 1.0, 0.9, 0.8, 0.1\}$. The normalized similarities

Characteristic	Basecase 1	Basecase 2	Basecase 3	Basecase 4	Basecase 5
Bytes in	1.0	1.0	0.9994	0.9995	0.0151
Bytes out	1.0	1.0	0.9994	0.9506	0.0011
Connections in	1.0	1.0	0.9796	0.96	0.15
Connections out	1.0	1.0	1.0	1.0	0.1519
Packets in	1.0	1.0	0.9989	0.9978	0.1376
Packets out	1.0	1.0	0.9992	0.9967	0.1131
SYN-ONLY rate	1.0	1.0	0.9780	1.0	0.8903
SYN-ACK rate	1.0	1.0	1.0	1.0	1.0
Idle connection rate	1.0	1.0	1.0	1.0	0.0
Half-open connection rate	1.0	1.0	0.9934	1.0	0.8889
Service packets	1.0	1.0	0.8883	0.7190	0.0
Service bytes	1.0	1.0	0.8911	0.7150	0.0
Service connections	1.0	1.0	0.8940	0.8281	0.0
Source port packets	1.0	1.0	0.8941	0.7512	0.0
Source port bytes	1.0	1.0	0.8937	0.7499	0.0
Source port connections	1.0	1.0	0.8454	0.7459	0.0
TTL	1.0	1.0	0.7959	0.6401	0.5637
Mean	1.0	1.0	0.9475	0.8962	0.2445

Table 5.1: The normalized similarities of count and discrete characteristics for basecases 1 through 5.

Characteristic	wf
Bytes in	0.0434
Bytes out	0.0555
Connections in	0.057
Connections out	0.0504
Packets in	0.0499
Packets out	0.049
SYN-ONLY rate	0.0
SYN-ACK rate	0.0
Idle connection rate	0.0428
Half-open connection rate	0.2343
Service bytes	0.0887
Service connections	0.0586
Service packets	0.0868
Source port bytes	0.0801
Source port connections	0.0665
Source port packets	0.0798
TTL	0.2172
y-intercept	-0.26

Table 5.2: The weighting factors for the singular and discrete characteristics found with linear regression, using the goal similarities of {1.0, 1.0, 0.9, 0.8, 0.1}.

of the individual metrics vary as to how close they are to the expected values. As such, we use a linear regression model (as described in Appendix E) to find weights and a constant correction factor (y-intercept) to apply to the normalized similarities to scale and combine them to form the overall similarity value. The `setSimMatrix.pl` script, written by the author creates the input file for Weka’s (University of Waikato 2007) Linear Regression Model. Weka’s output is parsed by the `parseModel.pl` script, also written by the author; the weights that we found for our basecases are shown in Table 5.2.

Interestingly, most of the metrics have weightings in the range (0.04..0.09), however the half-open connection rate and TTL weightings jump up to the (0.21..0.24) range. The TTL should not be surprising, as this was something that we deliberately manipulated when we created the basecases. The half-open connection rate happened somewhat incidentally in that the modified connections in the basecases did not attempt to maintain proper ACK sequences. Consequently, most of the modified connections (which were proportional to the target goal similarity) ended up as “half-open”. This hopefully illustrates why the weightings must be calculated for any given instantiation of the methodology, as each environment is likely to have their own idiosyncrasies that will affect the metric weightings.

Once we have the metric weightings, we can find the scaled similarity just by summing

Basecase	Goal	Calculated
Basecase 1	1.0	1.0
Basecase 2	1.0	1.0
Basecase 3	0.9	0.9000
Basecase 4	0.8	0.8000
Basecase 5	0.1	0.1000

Table 5.3: The goal and calculated similarities for the first five basecases using the singular and discrete characteristics with the weights found by linear regression.

the scaled metrics:

$$cf + \sum_{i=1}^m value_i w f_i$$

where cf is the correction factor (y-intercept), m is the number of characteristics, $value_i$ is the normalized similarity for a given metric, and $w f_i$ is the weighting factor for that metric found by linear regression. The `calcSimLinear.pl` script, written by the author, does exactly this, and for our five basecases it produces the results shown in Table 5.3.

Given that we are now able to calculate a scaled similarity that matches our goal similarity, we will turn our attention to continuous metrics.

6

Connection Metrics

Numerous metrics, or features, have been used to describe TCP connections. These features are frequently of interest to the network intrusion detection community, and occasionally the network engineering community. While researchers may understand these terms when used by others, there are no formal definitions of these terms such that two researchers looking at the same connection will necessarily calculate the same values for all the metrics. This chapter will look at a number of specific ambiguous cases that illustrate the need for formal metrics. The proposed set of connection metrics, along with formal definitions of each are given in (Brugger 2007b). Next, we will look at issues of ordering (or lack thereof) by time in the trace file. We will conclude by looking briefly at creating logical connections for UDP and ICMP streams, allowing many of the same metrics to be applied to them.

6.1 Introduction

Researchers frequently need to describe TCP connections, using metrics from the widely accepted “Duration” to the less commonly used, yet still understandable “URG packet rate”, through more obscure measures such as the “Wrong packet resend rate”. The advanced network intrusion detection community, which typically uses machine learning to attempt to identify malicious or anomalous connections, will use these metrics as the machine learning features (or attributes) that their methods train and recall on (Axelsson 2000). These metrics are also used by the network engineering community; however, in this arena, researchers tend to focus on the effects of a single metric (such as the “Interpacket latency”), on a device or network stack, at a time (Leland et al. 1994; Paxson and Floyd 1995; Racz et al. 2003). As such, these metrics tend to be better defined.

It is important that researchers agree on the definitions of these terms. This is somewhat easier in the network engineering community, where an assumption is typically made that TCP connections will behave to specification, and that misbehavior is the result of correctable deficiencies in network stack implementations, or unforeseen interactions as networks scale in size. The lack of definitions is more problematic in the intrusion detection community, as attackers actively create connections that are outside protocol specifications in an attempt to break stack implementations, such as Christmas Tree Packets (Miller 2000). Since this misbehavior may occur on such a small scale, relative to the total amount of traffic on the network, a difference in definitions may cause one researcher to measure a value of 0.01 for a metric, while another researcher measures a value of 0.02 – double the measurement of the first researcher.

We will begin by considering the effect of the viewpoint on the metrics we are considering. Then we will look at the ambiguity in the term “TCP connection” itself. Next, we will review a selection of metrics that have been used in published works that are ambiguous, and how that ambiguity may cause problems when doing analysis. It is worth noting that none of the works cited offer any guidance with respect to the questions we raise. We will limit our focus to first-order metrics of the connections. While the intrusion detection community frequently uses second-order (or calculated) metrics, such as “The number of new connections in the last n seconds” (Lee and Stolfo 1998; Lee et al. 1999a; Lee et al. 1999b; Lee and Stolfo 2000; Lee et al. 2000), we will not delve into those, as any ambiguity from those follows from the first-order metrics we do examine. Finally, we will encourage the reader to consider the formal definitions of TCP connection metrics that we have assembled, along with libpcap-based C code to capture them from pcap files.

6.2 Viewpoint

It turns out that the observer’s view of the network is rather important when considering TCP connection metrics. There are three general views one can have of network connections: from an endpoint, an active third-party observer, or a passive third-party observer. The reason this is important is that no single observation point has a comprehensive view of the connection. This section will look briefly at how each of these viewpoints differs from the others.

An endpoint is the source or destination host of a connection. The actions of an endpoint in response to any given TCP packet, given the current state of that connection, are well defined in RFC 793 (Postal 1981). An observer at this endpoint has the advantage of knowing what this state information is. While this is an important advantage, the major disadvantages come from coverage and complexity. A single endpoint can only see the connections to or from that one

host (including broadcast traffic), hence, in order to get broad network coverage, a great number of hosts may need to be instrumented. Besides the additional effort this requires, it is made harder by the second disadvantage: the complexity of instrumenting all the different operating systems that may appear on a given network. Certainly, the difficulty in instrumenting any given OS will vary depending on availability of the source, hooks provided for such instrumentation, and modularity of the code. However, even if we are given a kernel module (or equivalent) to do such instrumentation, it will likely not be acceptable to install it on many carefully managed production systems.

A third-party active observer is one that sits between the two endpoints of the connection, and has some control over that connection, for instance a firewall or intrusion prevention system. Such a viewpoint certainly has better network coverage, and likely has useful state information as well, for example, most firewalls keep the state of current connections, and will only admit a packet if it is establishing a new connection, or part of an established connection. Most such devices also have the ability to log or otherwise output information about the connections they allow or deny. Unfortunately, the format of such logs varies widely from device to device, as does the level of detail provided about the connections. Some of the metrics we will consider below (such as the resend rate), are not output by any firewalls or IPS devices that we are familiar with. Again, we could likely address the log format and metrics collected with internal access to the devices, but this is even less likely to happen, given the strict configuration management and device approval procedures required for such security sensitive equipment on many networks.

This leaves us with the third-party passive observer viewpoint. The advantage of this viewpoint is that it gives us the potential for coverage that the third-party active observer does, without the complications of tying into the internals of the device. It also allows for observation of networks without such devices (say, from an edge router), and allows us to use standard tools such as tcpdump (TCPDUMP Project 2006) to collect data. For these reasons, this is our preferred approach, and the viewpoint that we will use for the rest of this work. It is not without its disadvantages though:

1. We do not have the endpoint state, so even if we think a connection is established, if a host is rebooted, the connection may either suddenly be reset, or silently close.
2. We might miss packets because the observing machine can not keep up with the traffic on the wire.
3. Our observation may begin after some connections are already established, and some connections may continue after our observation ends.

Regardless of which of these techniques is used, none has a comprehensive viewpoint of all the connections on a given network. While either active or passive third-party might see a packet that gets dropped before making it to the end host, it is equally likely that such a packet drop occurs before even making it to the observation point. It is also worth noting that our primary interest is in internetwork connections (those that transit to or from the observed network). While there is no reason that a third-party observer could not watch the intranetwork connections through the trunking port of a switch, the engineering costs of capturing that much traffic are likely to be significant on contemporary high-speed networks.

6.3 Ambiguity in defining TCP connections

If we are interested in metrics on TCP connections, we must first have a good idea of what exactly a TCP connection is. For the purposes of this section, we will consider a TCP connection to just be the collection of packets that make up a single flow. While this seems easy by following (Postal 1981) while observing a well behaved connection, many ambiguities result in the edge-cases.

The first ambiguity comes primarily from our viewpoint: what if we did not start observing the link until after the connection started, or what if we stop observing the link before the connection ends? While our viewpoint may exaggerate this problem, it does not come solely from being a third-party observer. While a first-party observer may be able to utilize its host's TCP stack to determine if the host has initiated or accepted a new connection, what do we do with non-SYN packets that are not a part of an established connection? The host will just drop them, however we may want to collect metrics from them, as they may be part of some sort of nefarious activity, such as host fingerprinting. The question of when a connection ends is even more problematic, even for first-party passive observers: if many minutes have gone by since the last packet was observed in the connection, the observing host may consider the connection still open while the other host may have gone down (a half-open connection), which implicitly ends the connection, as the first host will discover if it attempts to send another packet.

Once we have a established connection, we will see packets that may or may not be part of the connection. For example, we may get the SYN flag set in an established connection – an obvious violation of protocol that should result in a RST. Is the illegal SYN packet part of the connection? What about packets with incorrect checksums? The destination will drop them and expect them to be resent. Do we count them as part of the connection? A couple related tricks frequently used to intrude onto a connection are sequence injection attacks and session hijacking. The first may be used as part of a social engineering attack by replacing some information in

the connection with what the attacker wants the other party to see. The second attack is where an attacker tries to usurp a connection – for example a telnet connection after the real source has authenticated to the destination machine. For either of these attacks, if the attacker can not stop the flow of packets from the real source, the destination will end up seeing some repeated segments; if the attack fails, the destination will see a number of out of sequence packets. Either way, do we count these packets – which the destination host drops – as part of the connection?

What about when an error ensues? If an uncorrectable error occurs, the host is supposed to issue a RST to close the connection. Do we count that RST as part of the connection? What if instead of a RST, we get an ICMP error message, such as “Host not available”? On the subject of errors, what about connections that never make it to an ESTABLISHED state? Do we even count them as connections?

The purpose of this chapter is to raise questions, but we should only worry about addressing these questions if they are relevant. Agreeing what packets are and are not part of a connection affects all other metrics about that connection. One of the most basic metrics is the packet count in the connection. This may be a small difference – for instance a difference of one if observer A counts a RST packet as part of a connection and observer B does not. It can also be a big error though – for instance if the sender sends hundreds of more packets after the receiver sends a RST. These differences propagate to all of the other metrics, for example the count of the number of errors on a connection; if observer A considers the connection closed after the first error, they will never get an error count of more than one, but if multiple packets get sent that each illicit a RST, observer B may count many more errors. Obviously, they will not be able to agree on anything meaningful about the network if observer A says they only saw one error and observer B says that they saw hundreds.

6.4 Ambiguous TCP connection metrics

There are numerous metrics one might use to characterize a TCP connection, for example the basic “Number of packets” measure discussed in the previous section. In this section we will discuss a number of connection metrics which have been used by numerous researchers in the network intrusion detection community, how they might be interpreted differently by different researchers, and what effect this difference in interpretation might cause.

6.4.1 Number of data or control packets

Besides the total number of packets in a connection, sometimes we are interested in the number of data or control packets in a connection. This metric has been cited in Lee and Stolfo (1998), Lee (1999), and Singh and Kandula (2001).

Is a packet either a data packet or a control packet? Put another way, can we calculate either metric by subtracting the other from the total number of packets? Or might a packet be both? Since all the packets after the initial SYN should ACK a segment from the other host, in a manner all packets should be control packets. Certainly, the difference between these two interpretations will create a large difference in the calculated values we come up for the metrics. Why would anyone even use the second interpretation if it is always going to report 100% of the packets are control packets? Well, if the number of control packets is less than the total number of packets, it indicates that something odd occurred in the connection, such as someone attempting to exploit a flaw in the other host's TCP/IP stack. If an intrusion detection system is built with (or learns) this rule, then it is going to have a serious problem if use the other interpretation and tell it that only 50 of the 100 packets we counted in a given connection were control packets.

These metrics can have more subtle differences as well: is an ACK packet that does not include any data (that is, the length of the data portion of the packet is zero) a data packet, a control packet, or neither? Or, taking the opposite approach of above, maybe one system only counts the packets with the SYN, RST, or FIN flags as control packets, in which case we will expect to see a control packet count of four (SYN, SYN-ACK, and two FINs) for a normal connection. That system is going to think something is seriously wrong if it is given some connection metrics that record 50% of the packets were control packets, like each of the four control packets had to be resent about a dozen times – maybe not an attack, but certainly something about the network that should raise some red flags.

6.4.2 Resend rate

The resend rate is a common measure of the reliability of a network. Lee and Stolfo (1998), Lee (1999), Chittur (2001), and Singh and Kandula (2001) all used it as one of the metrics feeding their network intrusion detection systems as well. We contend that the resend count is a slightly more useful metric, as the rate can easily be found by dividing by the total number of packets in the connection¹, and we do not have to worry about floating point error if we want to

¹At least we presume that these authors calculated the resend rate as the number of resends versus the total number of packets; one might also use the number of unique segments that were resent versus the total number of packets, this presents another ambiguity for consideration.

calculate more complex metrics, such as the resend rate over the past n connections. But how exactly do we count the number of resent packets in a connection?

For starters, we have a natural limitation from our viewpoint as a third-party observer. Ideally, we could observe a connection from both of the end hosts, in which case we could easily count how many times each of the hosts had to resend a given packet. Since we rarely, if ever, have this level of access, we must consider what we see monitoring the link between the hosts.

Do we only count packets as resent if we see the same packet multiple times? Or do we count the number of packets whose sequence numbers are less than the highest sequence number seen so far? That might be easy to do (and, in fact, is the approach used by the Ethereal network protocol analyzer(Ethereal, Inc. 2007)). Perhaps one should count the number of packets whose sequence numbers are less than the next logical sequence number we expect to see from the sender?

None of these solutions are foolproof given our position as a third party observer: keeping track of which packets have already been seen does not address packets that were lost between the sender and the observer. Counting packets with a sequence number less than the highest sequence number seen so far will count out-of-order packets as resends. Comparing instead against the next logical sequence number will only detect missing packets if we see a packet that follows it.

While no definition is perfect, what is more important is using a consistent definition. If one is attempting to compare the reliability of two networks, which in practice have the same average resend rate per connection, but one resend count is calculated by the first definition by an observer who does not see most of the duplicate packets, and the other is calculated by the third definition on data intensive connections (like file transfers) that allow it to easily see when packets were resent, will falsely show the second network to have a much higher resend rate (and hence, lower reliability) than the first network.

6.4.3 Number of establishment errors

One of the more popular, and most vaguely defined, connection metrics used by the network intrusion detection community is the number of establishment errors. This is used by Singh and Kandula (2001), Lee and Stolfo (2000), Lee et al. (2000), Lee and Stolfo (1998), Lee (1999), Chittur (2001), Lee et al. (1999b), Portnoy et al. (2001), and Dickerson and Dickerson (2000).

What exactly is an establishment error? Is it just a connection where we see a SYN packet, but that never makes it to the ESTABLISHED state? Can a connection that makes it to the ESTABLISHED state be considered to have had an establishment error? Can a connection

have more than one establishment error? Can a connection have an establishment error after it reaches the ESTABLISHED state? Most of the work cited above was interested in the number of establishment errors over the past n connections or x seconds, so the use of this metric does not shed that much light on these questions. Some of these connections hearken back to the question of exactly what is our definition for a TCP connection? If we do not consider something to be a connection unless it reaches the ESTABLISHED state, then that limits how we can interpret this metric.

With this large number of questions, it is easy to see how different definitions will cause confusion. If one researcher counts a single establishment error if they see a connection that does not make it to the ESTABLISHED state, and a different researcher counts the number of errors (bad checksums, incorrect TCP flags, exceeding window sizes, etc) on connections before they reach the ESTABLISHED state, they are comparing two entirely different things.

6.5 Sorting by time

Now that we have established formal definitions for our connection metrics and extended the `decodeTcpdump` program to output them. Immediately we recognized some problems, such as negative connection durations. Unfortunately, we had not found the secret to time travel – we merely had packets out of order in the pcap dump. The way we were sorting the data was essentially using a libpcap-based binary as the inner loop of a merge sort, with a Bash shell running the outer loop. The binary reported how many records it moved, and the loop would continue until bash saw that zero rows had been moved. The problem with this approach is that program return values in UNIX only use a single byte, so we only saw the modulo 256 of records that were moved, and stopped when that value reached zero.

The solution to this for now is to read all the records into a big array and use the `mergesort(3)` function to sort them. This was done using the `sortTime` program, written by the author. This works as long as the pcap dump fits into memory. We also created the `checkTime` program, which prints out any packets that appear out of order. We are hoping that when we get real data from large networks, that it is actually in order. If not, we will likely have to split it up by time into smaller segments, sort those individually, and merge them back together.

6.6 Logical Connections

We consider some ICMP activity as “logical connections”. While ICMP is a connectionless protocol, some activity should only be seen in response to other ICMP activity. Specifically,

an ICMP echo (type 8), timestamp (type 13), or information request (type 15), should elicit an ICMP echo reply (type 0), timestamp reply (type 14), or information reply (type 16), respectively. When we see one of these request types, we open a connection record, which we close and provide connection metrics on when we see the corresponding response, with a matching ID and sequence number. If we do not see both the request and reply, we provide only the packet-based metrics. Currently, we only match ICMP responses to ICMP requests. One potential area for improvement is to match ICMP responses to TCP requests, for instance ICMP destination unreachable packets in response to TCP SYN requests.

UDP is also a connectionless protocol; however, unlike ICMP, applications that run on top of UDP can have extremely rich protocol semantics, such as remote file systems. We expect that by capturing “logical connections”, where we group all the packets between the same pair of hosts and UDP ports, we can utilize the same metrics for UDP “connections” as we do for TCP connections, allowing us to better characterize the network. When a UDP packet comes in from a host/port set we have not seen before, we create a connection record for it. As each UDP packet from the same host/port set comes in, we see if more than TIMEOUT seconds (300 seconds, as above) have elapsed since the last packet; if so, we print out the statistics on the last “logical connection”, and start a new one; otherwise, we update the connection record. At the end of the pcap dump file, we print the connection metrics for all the UDP logical connections which had not seen a packet within TIMEOUT seconds of the last packet in the pcap dump file. This behavior is configurable using a flag on the `decodeTcpdump` program, allowing us to compare performance both using, and not using, the UDP logical connections.

There is one side-effect of ICMP and UDP logical connections, which should be disclosed: one of the metrics is the connection interarrival time. This is the delta, in seconds, since the last connection was started. This will include the above discussed ICMP requests, and UDP logical connections, if enabled.

7

Initial Continuous Characteristic Tests

Continuous characteristics are those that we can somehow plot out and fit a curve to. We expect that by taking the curves for the measured continuous characteristics for two traces, we should be able to compare the two curves and assign a similarity value. These are by far the trickiest measurements to characterize, because we will need to determine which function to assign to the characteristic to map the measurements to a curve. Our preference will be to start simple, with a second-degree quadratic equation, which we can map to using a least-squares fit. Failing that, there are almost a limitless (no pun intended) number of functions we can try, starting with higher order quadratics, principle component (PC)¹ or Fourier analysis, fractal distribution, wavelet modeling, and so forth. Likely, any function will be an approximation: Gudkov and Johnson (2001) indicate that most TCP/IP metrics can be described by an equation of 10-12 dimensions (or parameters), however they note that the form of this equation is an open problem.

Once we have assigned an equation to the continuous metric and determined the particular parameters for each trace in a pair, then we need to compare the two functions. With something simple like a quadratic equation, this will probably be as simple as comparing the samples of the curves at regular points. Other types of functions will likely be more complicated, for instance we can likely compare the Fourier analysis of two curves by calculating a weighted

¹It should be noted that while Lakhina et al. (2004) had some initial success in applying PCA to at least one network characteristic, Ringberg et al. (2007) found numerous limitations in the use of PCA on network traffic, particularly when applied for anomaly detection; even in our domain of network comparison, their cautions of sensitivity and scalability would be well heeded.

average of the similarities of the component functions. For something like a fractal distribution, we really have no idea at the outset how a comparison would be done.

Once we have all our characteristics modeled and comparable, we can find their normalized and scaled similarities and see how well they all work on the first five basecases.

7.1 Measuring

Another thing that makes the continuous characteristics harder than the others is that there are just so many of them. For starters, we have TTL and connection source port, which we already modeled as discrete characteristics, however they might be better modeled as continuous functions. We previously considered modeling some other characteristics, such as the per host packet, byte, and connection counts as continuous functions, however when we looked at them to determine what the independent (x) and dependent (y) variables would be, it became apparent that those metrics were not suited to be modeled as continuous distributions.

Despite the characteristics that are not suited to being modeled continuously, we still have 83 that are. These are enumerated in appendix B.

Now that we have formally defined our connection metrics (see chapter 6), we can extend the `decodeTcpdump` program to spit these out, along with any other metrics we need to build the distributions. Some of the connection metrics have been defined to be concrete, such as the number of fragmented packets, instead of the fragmentation rate, as the rate is represented as a floating point number. This floating point number is subject to round off error, unlike the count, so our formal metrics specify using the count, such that we can be assured that all researchers agree on the same value. Besides, the rate can be calculated from the counts, however not necessarily the other way around. Because of this, we run the output of `decodeTcpdump` through our `buildMnwRates.pl` and `buildConnRates.pl` scripts to generate rate values for the metrics that specify them.

We can continue to use the `buildDistro.pl` program to collect the measurements for each metric. Now for each continuous metric we need to define the function we can use to model its distribution.

7.2 Modeling continuous distributions

Here will will enumerate our many metrics that we expect to follow a continuous distribution, and specify what function we will use to model the distribution. For now, we will use

just the original one hour of data common to all five of our initial base cases, to determine the function to use.

Each continuous characteristic was specified to be either ordered or sorted. Ordered characteristics are those that we have both a X and a Y value for, where the X (independent variable) may be something like the TTL value, or minutes after the hour, and the Y (dependent variable) value represents the corresponding count. Most of the ordered continuous characteristics have a defined range, for instance TTL values can only be between zero and 255, inclusive; minutes after the hour can only be between zero and 59, inclusive.

Sorted continuous characteristics consist of a collection of measurements, such as the interlatency time between packets. We could plot the measurements across the X-axis and the count of packets or connections that matched that exact measurement as the Y-value, however many of the measurements are floating point values, and trying to fit a function through points where there may be a significant difference in Y-value between X-values that differ by a tiny (10^{-5}) amount will not work well; this makes sense considering that the measurements themselves are dependent variables. We have hence chosen the approach of sorting the values and plotting them successively. This should give us a nice continuously incrementing curve to fit a function to. Our intuition is that if one wanted to simulate the given characteristic, they would need only select a value at random from this curve. This assumption should be verified with empirical research at some point; we leave it as an open topic for future work.

These are the two useful ways that we have found to handle continuous characteristics. Another approach that has been employed is to plot the values successively over time, as was done with interconnection latency in (Paxson and Floyd 1995). This is certainly an option, and nothing in the methodology precludes it. In fact, the methodology encourages experimentation with different methods to find the one that best fits the data for a given characteristic – our choice to not look at time ordered data at this point is primarily due to the added complexity of such methods and the desire to complete the proof of concept of the methodology in a reasonable amount of time. This leaves the possibility of evaluating those characteristics in a time ordered manner for future work.

Originally, this section contained an in-depth analysis of all 83 continuous characteristics. This analysis was so comprehensive that it almost doubled the length of the main body of this thesis. As such, we have pulled this analysis out into appendix H, with the specific analysis on Interpacket Delta pulled into appendix F and TTL pulled into appendix G². We leave the definitions for these three appendices here, followed by a summary of the three methods we have

²Both Interpacket Delta and TTL have much more extensive analyses which serve as the basis for the analyses on the other characteristics.

determined are best to model our 83 continuous characteristics. A summary of the mappings between the characteristics and the modeling methods determined by the analysis in appendix H is given in Table 7.1. While this may seem like a lot of characteristics, they will be processed in bulk and subsequently assigned weights or dropped from the similarity calculation, as dictated by our tests on the base cases.

Table 7.1: Characteristics of TCP/UDP/ICMP/IPv4 traffic.

Methods used for modeling and comparison are count and ratio (for single value), discrete, and sorted, ordered, and non-keyed sorted for the continuous characteristics.

Single value	
Packets in (count)	Packets out (count)
Connections in (count)	Connections out (count)
Bytes in (count)	Bytes out (count)
SYN-ONLY rate (ratio)	SYN-ACK rate (ratio)
Idle connection rate (ratio)	Half-open connection rate (ratio)
Discrete	
Packet Service	Bytes Service
Connection Service	Packet Source port
Bytes Source port	
Connection Source port (also ordered continuous)	Packet TTL (also ordered continuous)
Continuous, per packet	
InterPacket delta (sorted)	Packet sec (ordered)
Packet min (ordered)	Packet GmHour (ordered)
Packet LocHour (ordered)	Packet weekday (ordered)
Bytes sec (ordered)	Bytes min (ordered)
Bytes GmHour (ordered)	Bytes LocHour (ordered)
Bytes weekday (ordered)	Packet size (ordered)
Continuous, per packet, time rates	
Packets in last w secs (ordered)	Priv packets time rate (sorted)
Unpriv packets time rate (sorted)	Connections time rate (sorted)
Priv connections connection time rate (sorted)	Unpriv connections connection time rate (sorted)
Continued on next page	

Table 7.1 – continued from previous page

Priv packets priv connection time rate (sorted)	Unpriv packets unpriv connection time rate (sorted)
SYNs connection time rate (sorted)	RSTs connection time rate (sorted)
FINs connection time rate (sorted)	PSH connection time rate (sorted)
Establishment errors connection time rate (sorted)	Other errors connection time rate (sorted)
Disconnection errors connection time rate (sorted)	Ave duration over last w secs (sorted)
Continuous, per packet, packet rates	
Priv packets packet rate (sorted)	Unpriv packets packet rate (sorted)
Continuous, per connection initiation	
InterConnection delta (sorted)	Connection sec (ordered)
Connection min (ordered)	Connection GmHour (ordered)
Connection LocHour (ordered)	Connection weekday (ordered)
Continuous, per connection initiation, connection rates	
Connection packet rate (sorted)	Connection Priv connections rate (sorted)
Connection Unpriv connections rate (sorted)	Connection Priv packet rate (sorted)
Connection Unpriv packet rate (sorted)	Connection SYNs rate (sorted)
Connection RSTs rate (sorted)	Connection FINs rate (sorted)
Connection PSH rate (sorted)	Connection Establishment errors rate (sorted)
Connection Other errors rate (sorted)	Connection Disconnection errors rate (sorted)
Ave duration over last m connections (sorted)	
Continuous, per connection close	
Number of packets (ordered)	Number of packets in (ordered)
Number of packets out (ordered)	Duration (sorted)
Number control packets rate (sorted)	Number data packets rate (sorted)
Number bytes transferred (ordered)	Number bytes transferred in (ordered)
Number bytes transferred out (ordered)	Number data bytes transferred (ordered)
Number data bytes transferred in (ordered)	Number data bytes transferred out (ordered)
Fragmented packets rate (sorted)	Bad fragment rate (sorted)
Continuous, per connection close, TCP only	
Max Src Window (ordered)	Max Dst Window (ordered)
Urgent rate (sorted)	Resend rate (sorted)
Wrong resend rate (sorted)	Duplicate ACK rate (sorted)
Wrong ACK (sorted)	Wrong data packet size rate (sorted)
Continued on next page	

Table 7.1 – continued from previous page

Window exceeded rate (sorted)	Hole rate (sorted)
Number connection errors (ordered)	Number reset connection (ordered)
Number other errors (ordered)	Number disconnection errors (ordered)
Packet Destination IP (non-keyed sorted)	Bytes Destination IP (non-keyed sorted)
Connection Destination IP (non-keyed sorted)	Packet Source IP (non-keyed sorted)
Bytes Source IP (non-keyed sorted)	Connection Source IP (non-keyed sorted)

7.2.1 Definitions

The remainder of this section describes the modeling methods for (TCP/UDP/ICMP)/IPv4 traffic, with discussion of the particular characteristics pulled into appendices as described above. Across this work, we use some terms which we will define here to avoid repeating ourselves.

Definition. **Connection rate** is a measurement made over the past m connections. If the number of connections that we have information on is less than m , we use that smaller number as the denominator for calculating our rates.

Definition. **Packet rate** is a measurement made over the past n packets. If the number of connections that we have information on is less than n , we use that smaller number as the denominator for calculating our rates.

Definition. **Privileged connection** is a TCP or UDP connection where the destination port of the connection is < 1024 . Some connections, such as ICMP, are neither privileged nor unprivileged.

Definition. **Privileged packet** is a packet that is part of a *privileged connection*, regardless of which way it is going.

Definition. **Time window** is the number of past seconds – referred to as w – over which a measurement is made.

Definition. **Unprivileged connection** is a TCP or UDP connection where the destination port of the connection is ≥ 1024 . Some connections, such as ICMP, are neither privileged nor unprivileged.

Definition. **Unprivileged packet** is a packet that is part of a *unprivileged connection*, regardless of which way it is going.

We use a value of $m = 50$ for the connection rate, which is somewhat arbitrary (that is to say, an educated guess), given the lack of research into good choices for this parameter; this would be a good topic for future research. In fact, since we had less than 50 connections in our sample data, our choice is almost irrelevant (almost, as it could have been relevant if we made it smaller); however, it will matter later in our tests with real data.

We use a value of $n = 50$ for the packet rate, which is somewhat arbitrary (that is to say, an educated guess), given the lack of research into good choices for this parameter; this would be a good topic for future research.

The choice of values for the time window, w , can range from very short (such as $w = 3$) to very long ($w = 86400$ or one day) (Lee and Stolfo 1998; Lee, Stolfo, and Mok 2000; Barbará, Wu, and Jajodia 2001). We use a fairly popular intermediate value of $w = 30$. It is an open question for future research how the choice of this value affects the metrics, and what the ideal value is for any given metric. A connection is considered active in this time window if at least one packet is transmitted in either direction.

7.2.2 Modeling ordered continuous characteristics

Our analysis of modeling TTL values as a continuous characteristic is detailed in appendix G. From the start, we recognized that the order of the independent variables (the actual TTL values) was important, so we plotted out the data with these TTL values along the X-axis and the count of packets with each of these values as the Y-axis, as shown in Figure G.1. We attempted to model this characteristic as a polynomial, or using a Fourier transform.

Ultimately we determined that it was best to just save the points, as we do with discrete characteristics. The difference from discrete characteristics is that – in order to take into account the continuous nature of the data, we will actually compare a weighted average of each point, consisting of the value at that point, plus half the sum of the adjacent points, plus a quarter of the points adjacent to those, divided by 1.75.

7.2.3 Modeling sorted continuous characteristics

Our analysis of modeling the interarrival time of packets as a sorted characteristic is detailed in appendix F. When we started looking at the interpacket delta, we recognized that – unlike TTL – there was not an independent variable, unless we considered the time of arrival. Just taking all the values and plotting them in sorted order produced a nice curve, as shown in Figure F.1. Despite a wealth of prior work indicating this curve followed a power-law function, we could not get a very tight fit of the data to such a function.

Eventually we determined that such a model would not buy us anything in the comparison phase anyway. Once again we elected to just use the sorted data points as our model and perform our comparison between two sets of data by projecting the larger dataset to the domain of the smaller dataset and find the mean difference between each point on the smaller curve and a projected point in the same x position on the larger curve.

7.2.4 Modeling non-keyed sorted continuous characteristics

Our third continuous characteristic model actually came from the analysis we did in chapter 5 specifically concerning IP addresses. It is actually a combination of the two above approaches: the actual modeling is done using a map, as with discrete or ordered characteristics. The keys to this map are then thrown away and the values are then sorted and compared as with sorted characteristics. The beginning of section H.1.7 contains additional details.

7.3 Normalized Similarity

With all of our characteristics now defined, we updated the `normCompare.pl` script with the three comparison methods:

1. Comparing the weighted average of every point between two plots ordered by value (ordered characteristics)
2. Comparing every point on the curve with less points to a (possibly projected) point on the curve with more points, where both curves are sorted by their key values (each of which may occur multiple times) from lowest to highest (sorted characteristics)
3. A variant on the previous one that ignores the key values and just uses sorted plots of the counts (non-keyed sorted characteristics)

With this done, we can run `normCompare.pl` on each of our basecases, producing the results shown in Table H.1. Note that we also include our singular and discrete metrics here, as we will be using them in the next step – the values for them should be redundant with what was shown back in Table 5.1.

The only thing unusual about these similarities that immediately jumps out at us is that the second bas case has some values that are less than 1.0. The second bas case, you may recall, is testing that the same trace with the timestamps advanced by one hour is considered to

be equivalent to the first trace. In other words, if we want to see how a network has changed over time, and it has not, we should get a similarity of 1.0. As you will see, this is corrected when we look at the scaled similarities.

The only other observation we can make at this point is that we did not really do a good case of making the pairs in baselines three, four, or five, as different as we wanted, with respect to many of the metrics. This is fine, as the purpose of the baselines was primarily to test an initial instantiation of the methodology. What will be much more informative is how these similarities look on real network traffic.

7.4 Scaled Similarity

When we run a linear regression on our normalized similarities shown in Table H.1, we get the weighting factors shown in Table H.2. This gives us pause, as some of the weights are not only greater than 1.0, but significantly so, and the constant correction factor (y-intercept) has also grown by over two orders of magnitude. On the other hand, all the measurements with local or UTC hour, which resulted in normalized similarities of less than 1.0 for the second baseline, have been set to zero. If we go forward and run the baselines through the calcSimLinear.pl script with these weights, we get the calculated similarities shown in Table 7.2.

Baseline	Goal	Calculated
Baseline 1	1.0	0.9996
Baseline 2	1.0	0.9996
Baseline 3	0.9	0.8996
Baseline 4	0.8	0.7997
Baseline 5	0.1	0.0999

Table 7.2: The goal and calculated similarities for the first five baselines using all characteristics.

A known limitation of linear regression is that it attempts to use all the data available to fit its lines, which can result in over-fitting. Out of curiosity, we took the top 10 weighted features and tried rerunning the linear regression with just those features. The weights ranged from 0.4159 to 181.036 with a y-intercept of -603.5161, and performance was essentially unchanged. This indicates to us that attempting to reduce the features available to linear regression is actually increasing over-fitting.

It is hard to argue that the approach as it currently stands is giving us incorrect results, given the results shown in Table 7.2, but then we are using the same five baselines for both training and testing. The real test, it would seem, is when we try to apply this to real data and have separate training and test cases.

8

Tests on real data

Given the success we have had in reconstructing our desired similarities on artificial data, it is time to try applying the methodology to real network data. This chapter will begin by looking at our available data sources, how we will divide them up into baselines, both for training and testing, the normalized similarities of all the metrics on those baselines, the weighting factors based on the training cases, and the scaled similarities of the test cases.

8.1 Usable network traces

Appendix I shows all the notes we made as we searched for suitable network traces for our tests. This section only summarizes the relevant data regarding the data sources that we have decided to use, with notes about what scenarios they should be useful for and technical limitations in their use. We present these datasets – one per subsection – roughly in order of usefulness for our tests.

8.1.1 Linode

Description: The author has a Linux virtual host collocated server, operated by [Linode.com](#). Numerous packet header traces have been captured from it spanning months over the past couple years.

Topology / Policy: The network consists of a single host in a CoLo facility a couple hops from an Internet backbone router. The host is only used by a single user (the author) for SMTP mail and other assorted uses as the need arises.

Limitations:

- Despite the large amount of data captured, very little of it intersects with other available network captures
- Very limited topology
- No other CoLo networks for comparison

Usefulness:

- Months worth of data spanning a couple years
- Author knows what all the authorized activity on this network was
- Includes a lot of external activity, including attacks and network noise
- Extremely useful for comparing how a single network changes over time

8.1.2 DSL1

Description: The author has captured packet header traces on his residential DSL line.

Topology / Policy: The network consists of multiple nodes (wired and wireless) in a residential setting connected to a Linksys router, providing NAT functionality, converting the traffic from the multiple nodes to use a single IP. The captures take place between the NAT router and the DSL modem, hence it captures both the outgoing traffic (all of which is presumably authorized), and incoming traffic, which includes attacks and network noise. The network consists of two users and three cats who were aware of the monitoring, which may have impacted their actions on the network. The Internet facing address of the network was assigned by DHCP and changed at times throughout the data collection. Some network traces were modified to make the IP address consistent throughout that trace.

Limitations:

- No other residential networks for comparison
- Users aware of study, which may have impacted their use
- The IP address of the monitored network may have changed over the course of the data collection

- Would be nice to compare against OSDI2006 traffic, collected at the same time; unfortunately, that data appears to have been randomized

Usefulness:

- Intersects with a couple other network traces, which it can be compared against
- Have weeks worth of data over multiple months, allowing us to evaluate how a network changes over time

8.1.3 Dartmouth campus

Description: Dartmouth collected packet header traces from an increasing number of access points of their campus wireless network over three semesters of multiple years. While this data was collected for the purposes of studying wireless network usage specifically, it appears very useful for general network usage studies as well (Dartmouth College 2008).

Topology / Policy: Packet traces are made at wireless access points at different buildings (and hence, departments) around campus. Each AP appears to use a well-defined network subnet for its clients. The policy is presumably open and academic in nature.

Limitations:

- Last data collects made four years ago and do not intersect with any other datasets

Usefulness:

- Collects were made for months at a time, so we can compare entire days worth of data from adjacent weeks
- Useful for comparing two different networks from different departments (and hence, expectedly different usage patterns)
- Useful for comparing how networks evolve over longer time periods (some of the networks have data spanning three years)
- While we do not have the explicit policies used, we know what building each trace was collected from, so we can infer the policy, users, and type of usage

8.1.4 LBNL

Description: ISIR at LBNL collected packet header traces covering most of all of the subnets at LBNL, and possibly related organizations, such as UCB, for a few minutes on three days across four months (Cooperative Association for Internet Data Analysis (CAIDA) 2008).

Topology / Policy: The traces appear to be made at a border router with 16 ports where each port provides service to one or more subnets. The policy is presumably a somewhat open research and development environment.

Limitations:

- Traces only last a few minutes each
- Traces from different ports on the same day may or may not overlap, and if they do, it is only for a couple minutes
- Do not know what the slight policy differences are between any two given subnets

Usefulness:

- Useful for comparing against a network with a very different policy (Linode) at the same time
- Useful for comparing how networks change over time
- Slightly useful for comparing two different networks at around the same time

8.1.5 SOTM27

Description: SOTM27 was a conference held in early March 2003. Packet header traces were made of the wireless network made available to the participants (Cooperative Association for Internet Data Analysis (CAIDA) 2008).

Topology / Policy: The capture was apparently made at the edge between the wireless network and the WAN. The policy was likely very permissive from the viewpoint of the participants. The capture indicates that the participants were given private IP addresses, and they likely shared a public IP via NAT. The fact that we see the private IP addresses indicates that the sensor is inside the router, and we should not see incoming traffic that was filtered by the router. The traffic we do see is very bursty in nature, with numerous gaps of more than five minutes between

adjacent packets; it is unknown if this is actually what the traffic looked like or if the sensor did not capture all the traffic.

Limitations:

- We apparently do not see incoming traffic – such as malicious connections, backscatter, and network noise – which is filtered by the router.

Usefulness:

- Have a few days of traffic, so we can compare the network over the time of day
- Useful for comparing to other conferences which presumably have similar policies

8.1.6 SIGCOMM2004

Description: SIGCOMM2004 was a conference held in late August, early September 2004. A wireless network was made available to participants, and packet header captures were made of both the wireless traffic, and the wired connection to the WAN (Dartmouth College 2008).

Topology / Policy: We are interested primarily in the traffic captured at the WAN interface. This shows two distinctly networks dominating the traffic, which we presume are the local networks anonymized. There appears to be some incoming traffic to this network, however it appears to be lower than we would expect given the levels of network attacks and noise observed even back in 2004, so the capture may have been made inside a firewall which allowed specified incoming traffic. The client policy was likely permissive from the viewpoint of the participants. The traffic shows very distinct patterns of daily usage, with gaps of sometimes hours between packets in the off-hours; however, we know that the sensors were working during these periods as they were capturing background noise such as ARP packets.

Limitations:

- Some amount of incoming traffic is apparently filtered

Usefulness:

- Have a few days of traffic, so we can compare the network over the time of day
- Useful for comparing to other conferences which presumably have similar policies
- Has a couple different networks, with different utilization levels, which could be compared

8.1.7 AMES Internet Exchange OC-48

Description: The AMES Internet Exchange was the original Metropolitan Area Exchange West (MAE-WEST), and continues to serve as a major peering point for commercial and government networks. Anonymized traces were made of one of the peering links for a few hours each over three days in 2002 and 2003 (Cooperative Association for Internet Data Analysis (CAIDA) 2008).

Topology / Policy: Traffic from numerous networks of differing topologies and unknown policies are captured. The traces we used came from the first collection covering three hours in 2002. This collection was missing the second hour of data from dag1, so we limited our scope to networks observed in the first hour on dag0 with less than 10% of the level of traffic on dag1 based on the idea that the missing dag1 data would then be inconsequential for the second hour. Similar and different networks were selected based on their levels of UDP and TCP traffic, number of initial SYNs observed, and levels of traffic to the top TCP services.

Limitations:

- Not sure we are capturing complete flow data (due to asynchronous routing)
- No idea what the policies are of the networks captured by the trace
- Does not overlap with any other datasets

Usefulness:

- Could use by extracting one subnet, splitting extraction into two halves and comparing those
- Could extract two different networks based on the presumption that the policies are significantly different if the traffic volume is significantly different (perhaps with human verification) and compare those
- Could extract one network from this dataset and a network from another peering point dataset – such as the Ampath data – which we can verify by hand to be different, and compare those

8.1.8 DefCon 10

Description: DefCon is the premiere hacker conference, held every summer in Las Vegas; DefCon 10 was held in early August 2002. It is an extremely interesting conference from a networks point of view, as an official conference wireless network is constructed for the participants use, as well as numerous rogue networks, attempting to entice participants to use their networks instead, for nefarious purposes. Most of the traffic is hostile, and one can be assured that any connections are intercepted, and any amount of malicious manipulation is being performed, including cache poisoning, connection hijacking, content injection, etc. One of the events that takes place at DefCon is the Capture the Flag contest, wherein multiple teams attempt to defend some designated tokens (the flags) on their servers, while capturing other teams flags by subverting the security on that team's servers. This event takes place on a dedicated network. For DefCon 10, Shmoo (a hacking organization), captured the data on the CTF network and made it publicly available (Shmoo 2007).

Topology / Policy: The network is a stand alone (separate from the Internet) network consisting of multiple subnets, with one subnet for each team participating in the contest. Each subnet was captured separately. The policy was completely open; participants were allowed to do anything over the network, and they did.

Limitations:

- Does not overlap temporally with any other dataset
- Extremely different from any other type of network described about

Usefulness:

- Can compare the different subnets to each other
- Could compare against the DefCon 9 CTF capture
- Useful as an exemplar of an “extremely different” network

8.2 Real basecases

Of the basecases covered in chapter 4, we need to construct the following from the above traces. The scores that we specify each set of basecases should achieve are based on our

expectation of the relative differences between all of the baselines and how those differences should map in our [0..1] range.

1. Two adjacent hours of traffic from the same network in mid-day (score should be > 0.75)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-08-06 18:00 64.5.53.67/32	Linode 2007-08-06 19:00 64.5.53.67/32	DSL1 2007-08-06 18:00 69.110.79.251/32	DSL1 2007-08-06 19:00 69.110.79.251/32
SOTM27 2003-03-02 18:00 172.16.134.0/24	SOTM27 2003-03-02 19:00 172.16.134.0/24	SIGCOMM2004 2004-08-31 19:00 98.93.251.0/24	SIGCOMM2004 2004-08-31 20:00 98.93.251.0/24
Ames net 2002-08-14 16:00 237.24.63.0/24	Ames net 2002-08-14 17:00 237.24.63.0/24	Ames net 2002-08-14 16:00 239.65.134.0/24	Ames net 2002-08-14 17:00 239.65.134.0/24

2. Two adjacent hours of traffic from the same network in off-hours (score should be > 0.75)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-08-06 09:00 64.5.53.67/32	Linode 2007-08-06 10:00 64.5.53.67/32	DSL1 2007-08-06 09:00 69.110.79.251/32	DSL1 2007-08-06 10:00 69.110.79.251/32
SOTM27 2003-03-02 09:00 172.16.134.0/24	SOTM27 2003-03-02 10:00 172.16.134.0/24	SIGCOMM2004 2004-09-01 09:00 98.93.251.0/24	SIGCOMM2004 2004-09-01 10:00 98.93.251.0/24
Dartmouth 2003-11-03 09:00 190.84.172.0/24	Dartmouth 2003-11-03 10:00 190.84.172.0/24	Dartmouth 2003-11-03 09:00 190.84.69.0/24	Dartmouth 2003-11-03 10:00 190.84.69.0/24

3. The same mid-day hour of traffic from the same network on adjacent weeks (score should be > 0.75)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-08-06 18:00 64.5.53.67/32	Linode 2007-08-13 18:00 64.5.53.67/32	DSL1 2007-08-06 18:00 69.110.79.251/32	DSL1 2007-08-13 18:00 69.110.79.251/32
Dartmouth 2003-11-03 18:00 190.84.172.0/24	Dartmouth 2003-11-10 18:00 190.84.172.0/24	LBNL 2004-12-16 21:17 128.3.23.0/24	LBNL 2005-01-07 23:26 128.3.23.0/24
Dartmouth 2003-11-03 18:00 190.84.69.0/24	Dartmouth 2003-11-10 18:00 190.84.69.0/24	LBNL 2004-12-16 21:17 131.243.155.0/24	LBNL 2005-01-06 22:25 131.243.155.0/24

4. The same off-hour of traffic from the same network on adjacent weeks (score should be > 0.75)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-08-06 10:00 64.5.53.67/32	Linode 2007-08-13 10:00 64.5.53.67/32	DSL1 2007-08-06 10:00 69.110.79.251/32	DSL1 2007-08-13 10:00 69.110.79.251/32
Dartmouth 2003-11-03 10:00 190.84.172.0/24	Dartmouth 2003-11-10 10:00 190.84.172.0/24	Dartmouth 2003-11-03 10:00 190.84.69.0/24	Dartmouth 2003-11-10 10:00 190.84.69.0/24
SOTM27 2003-03-02 10:00 172.16.134.0/24	SOTM27 2003-03-05 10:00 172.16.134.0/24	SIGCOMM2004 2004-09-01 10:00 98.93.251.0/24	SIGCOMM2004 2004-09-03 10:00 98.93.251.0/24

5. One mid-day hour and one off-hour of traffic from the same network (score should be lower than the above scores)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-08-06 10:00 64.5.53.67/32	Linode 2007-08-06 18:00 64.5.53.67/32	DSL1 2007-08-06 10:00 69.110.79.251/32	DSL1 2007-08-06 18:00 69.110.79.251/32
Dartmouth 2003-11-03 10:00 190.84.172.0/24	Dartmouth 2003-11-03 18:00 190.84.172.0/24	LBNL 2004-12-15 11:09 131.243.140.0/24	LBNL 2004-12-15 19:42 131.243.140.0/24
SOTM27 2003-03-02 10:00 172.16.134.0/24	SOTM27 2003-03-02 18:00 172.16.134.0/24	SIGCOMM2004 2004-09-01 10:00 98.93.251.0/24	SIGCOMM2004 2004-08-31 19:00 98.93.251.0/24

6. One hour of data from one network, and the same hour of data from a different network that is similar in size, topology, policy, user base, etc, such that we should expect the traffic to be similar in composition (score should be > 0.5 and < 0.75)

Train net 1	Train net 2	Test net 1	Test net 2
Ames net 2002-08-14 16:00 237.24.63.0/24	Ames net 2002-08-14 16:00 237.24.65.0/24	Ames net 2002-08-14 16:00 239.65.134.0/24	Ames net 2002-08-14 16:00 3.12.2.0/24
Dartmouth 2003-11-03 18:00 190.84.172.0/24	Dartmouth 2003-11-03 18:00 190.84.69.0/24	LBNL 2004-12-15 18:12 131.243.12.0/24	LBNL 2004-12-15 18:12 131.243.13.0/24
Dartmouth 2003-11-03 18:00 190.84.172.0/24	Dartmouth 2003-11-03 18:00 190.84.44.0/24	LBNL 2004-12-16 21:17 128.3.45.0/24	LBNL 2004-12-16 21:17 128.3.47.0/24

7. One hour of data from one network, and the same hour of data from a different network that differs in a single significant way (size, topology, policy, user base, etc) from the first such that we should expect the traffic to share some similarities, but still be less similar in composition to the above (score should be > 0.25 and $<$ above)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-08-06 18:00 64.5.53.67/32	DSL1 2007-08-06 18:00 69.110.79.251/32	Linode 2007-08-13 18:00 64.5.53.67/32	DSL1 2007-08-13 18:00 69.110.79.251/32
Dartmouth 2003-11-03 18:00 190.84.172.0/24	Dartmouth 2003-11-03 18:00 190.84.116.0/24	LBNL 2004-12-16 21:17 128.3.45.0/24	LBNL 2004-12-16 21:17 128.3.46.0/24
SOTM27 2003-03-02 18:00 172.16.134.0/24	SIGCOMM2004 2004-08-31 19:00 98.93.251.0/24	Ames net 2002-08-14 16:00 173.148.249.0/24	Ames net 2002-08-14 16:00 3.3.31.0/24

8. One hour of data from one network, and the same hour of data from a significantly different network (score should be < 0.25)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-11-05 18:00 64.5.53.67/32	Dartmouth 2003-11-03 18:00 190.84.172.0/24	DSL1 2007-02-01 18:00 69.110.74.139/32	Dartmouth 2004-01-29 18:00 190.84.172.0/24
Linode 2004-12-16 21:17 64.5.53.67/32	LBNL 2004-12-16 21:17 128.3.45.0/24	Linode 2005-01-07 23:26 64.5.53.67/32	LBNL 2005-01-07 23:26 128.3.23.0/24
Linode 2008-03-02 18:00 64.5.53.67/32	SOTM27 2003-03-02 18:00 172.16.134.0/24	DSL1 2007-03-02 18:00 69.225.88.159/32	SOTM27 2003-03-02 18:00 172.16.134.0/24

9. One day of data from a network, and one day of data from the same network on the same day of the following week (score should be > 0.75)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-08-06 00:00 64.5.53.67/32	Linode 2007-08-13 00:00 64.5.53.67/32	DSL1 2007-08-06 00:00 69.110.79.251/32	DSL1 2007-08-13 00:00 69.110.79.251/32
Dartmouth 2003-11-03 00:00 190.84.172.0/24	Dartmouth 2003-11-10 00:00 190.84.172.0/24	Dartmouth 2003-11-03 00:00 190.84.69.0/24	Dartmouth 2003-11-10 00:00 190.84.69.0/24
SOTM27 2003-03-02 00:00 172.16.134.0/24	SOTM27 2003-03-05 00:00 172.16.134.0/24	SIGCOMM2004 2004-08-31 14:00 98.93.251.0/24	SIGCOMM2004 2004-09-02 14:00 98.93.251.0/24

10. One day of data from one network, and the same day of data from a similar network (score should be > 0.5 and < 0.75) While we were able to generate three training pairs and three testing pairs for all the other baselines, the Dartmouth data was the only source of similar networks for which we had entire days of traces, hence we only use a single pair from this dataset for training, and a different single pair for testing.

Train net 1	Train net 2	Test net 1	Test net 2
Dartmouth 2003-11-03 00:00 190.84.172.0/24	Dartmouth 2003-11-03 00:00 190.84.69.0/24	Dartmouth 2003-11-03 00:00 190.84.172.0/24	Dartmouth 2003-11-03 00:00 190.84.44.0/24

11. One day of data from one network, and the same day of data from a slightly different network (score should be > 0.25 and < 0.5)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-08-06 00:00 64.5.53.67/32	DSL1 2007-08-06 00:00 69.110.79.251/32	SOTM27 2003-03-04 14:00 172.16.134.0/24	SIGCOMM2004 2004-08-31 14:00 98.93.251.0/24
Dartmouth 2003-11-03 00:00 190.84.172.0/24	Dartmouth 2003-11-03 00:00 190.84.116.0/24	Dartmouth 2003-11-03 00:00 190.84.172.0/24	Dartmouth 2003-11-03 00:00 190.84.224.0/24
Dartmouth 2003-11-03 00:00 190.84.172.0/24	SOTM27 2003-03-02 00:00 172.16.134.0/24	Dartmouth 2004-01-27 14:00 190.84.172.0/24	SIGCOMM2004 2004-08-31 14:00 98.93.251.0/24

12. One day of data from one network, and the same day of data from a very different network (score should be < 0.25)

Train net 1	Train net 2	Test net 1	Test net 2
Linode 2007-11-05 00:00 64.5.53.67/32	Dartmouth 2003-11-03 00:00 190.84.172.0/24	DSL1 2007-02-01 00:00 69.110.74.139/32	Dartmouth 2004-01-28 00:00 190.84.172.0/24
Linode 2008-03-02 00:00 64.5.53.67/32	SOTM27 2003-03-02 00:00 172.16.134.0/24	DSL1 2007-03-02 00:00 69.225.88.159/32	SOTM27 2003-03-02 00:00 172.16.134.0/24
Linode 2007-08-04 12:00 64.5.53.67/32	DefCon 10 2002-08-03 12:00 192.168.2.0/24	DSL1 2007-08-04 12:00 71.133.175.192/32	DefCon 10 2002-08-03 12:00 192.168.2.0/24

8.3 Normalized similarity

Altogether, we have 86 distinct traces.

We began by ensuring that each of the traces was sorted in temporal order. We proceeded to build the distribution of characteristics for each trace.

In the process we had to address memory management issues stemming from invalid states, such as out of order packets resulting in a reset (RST) packet after a connection reached the closed state. These fixes also addressed some incorrect metrics which had been reported by the processor. Most of the files could be processed given three gigabytes of memory. The two largest traces had to be processed one characteristic at a time using the `singleDistro` program with the `multiSingleDistro.pl` wrapper script.

The resulting distributions are shown in tables and figures in Appendix K. Just building this appendix was troublesome given the size of the distributions. We originally built the `genDistroGraphs.pl` script to do this. That script could not handle the largest characteristics of the large distributions, so we first presorted the sorted characteristics using the `sortChar` program, driven by the `sortDistro.pl` script. Insufficient, we pulled all the sorted continuous

characteristic processing into a compiled program called `buildSortedArray`, which builds the sampled array plotted by `gnuplot`. This allowed us to process the largest file by itself, but not all of the files together, even after moving most of our state storage off to disk; this is undoubtedly due to some opaque behavior of Perl memory management. Splitting the functionality into `singleDistroGraphs.pl` and `combineGraphs.pl` finally allowed us to generate the appendix.

Amazingly, once the distributions were built, actually doing the normalized comparison was much simpler. We did need to pull the comparison for sorted continuous characteristics out into compiled code, called `sortedCompare`, which is called by the `normCompare.pl` script used previously with some minor modifications, such as passing values (some of which are quite large) by reference. Interestingly, the `sortedCompare` program works correctly when compiled and run for the PowerPC architecture on Mac OS X, but not when compiled and run for the x86 architecture; this was tracked down to a scoping issue with variables in stdio loops by gcc, and has been reported to Apple. While examining this process, we found that some of the characteristics were not being scaled properly by connection, packet, or byte count, which was easily fixed by giving them names indicating which they required.

Table J.1 shows the intended goals for each of the pairs in all of the basecases, along with the calculated average (mean) of all the normalized similarities. We find very little correlation between these mean similarity values and our goals. This is to be expected. If there was a strong correlation, we would not need to assign individual weights and possibly prune the set of characteristics; put another way, if our calculations were giving us just what we expected at this stage, we'd be done. Since that is not the case, we will look at weighting and scaling the values in the next section. Appendix L presents the normalized similarities for each metric in all the training pairs.

8.4 Weighting factors and Scaled similarity

8.4.1 Training cases

For the artificial base cases, we had the advantage of knowing a priori what the goals for each of the training cases was supposed to be. For the real data, we are not as fortunate, and as Table J.1 shows, their normalized similarities are all over the map. Given this, we tried setting the goals three different ways: assigning a value within the goal range weighted based on the pair's normalized similarity, setting the same goal for all pairs of a given bas case, or setting the same goal, except for pairs with a similarity of zero, which are given a goal of zero. The selected goals are shown in Table J.2.

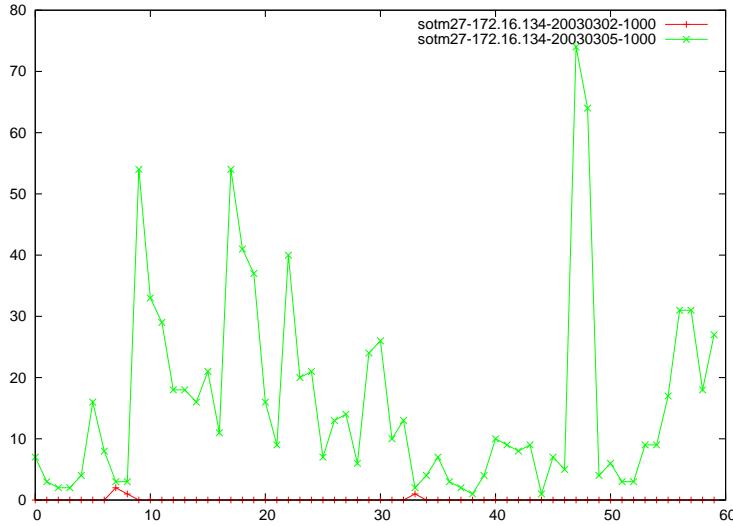


Figure 8.1: Number of connections for each second after the minute for basecase 4, pair 3

8.4.2 Validation

Now that we have built our weights, we need to validate that they are reasonable. To do this, we apply them back to our training cases, the results of which are shown in Table 8.1.

For the most part, the scaled similarities for the training pairs matches the goals we set for each. There are, however, a number which vary significantly (the difference is greater than 0.1) from the intended goal. Basecase 4, pair 3 is interesting because it is the only pair with a goal above 0.75 which was off the mark, as well as the only pair with scaled similarities outside our defined range of [0..1]. Both basecase 7, pair 3, and basecase 11, pair 3 were higher than their [0.25..0.5] targets. We see the greatest deviation at the low end, however, with all the pairs of basecase 8, as well as basecase 12, pair 2 all calculating higher than their goals, which were to be below 0.25. To help us understand this, we will take a look at a sampling of these pairs and see if our goals were simply off the mark.

For basecase 4, pair 3, we could show you all the plots of the distributions of all the characteristics for both traces, but instead we believe that Figure 8.1 does a great job of illustrating the differences between the two traces: the trace on March 2nd has only a small fraction of the connections that the trace on March 5th does. Not only are these two traces not as similar as planned: they are about as dissimilar as we would expect to get. So why did it score so highly? Because the first trace has so few data points, many of the normalized similarities are higher than we would expect, and these values likely resulted in increased weighting on those characteristics in order to achieve the unreasonable goal for this pair. There are two approaches we could take at this point: either drop the pair from the training set, or adjust its goal to put

Basecase	Pair	Scaled	Linear	Linear w/ 0s
1	1	0.83508	0.87497	0.87495
1	2	0.00000	0.87500	0.00000
1	3	0.85563	0.96355	0.95145
2	1	0.85510	0.87499	0.87497
2	2	0.00000	0.87500	0.00000
2	3	0.95008	0.87494	0.87489
3	1	0.84010	0.87499	0.87497
3	2	0.83010	0.87493	0.87490
3	3	0.92007	0.87486	0.87484
4	1	0.90010	0.87498	0.87491
4	2	0.82517	0.87497	0.87494
4	3	0.97092	1.05846	1.16443
5	1	0.70008	0.67498	0.67494
5	2	0.62506	0.67497	0.67505
5	3	0.72504	0.67499	0.67486
6	1	0.63070	0.66349	0.65128
6	2	0.52503	0.57493	0.57499
6	3	0.55009	0.57492	0.57492
7	1	0.39004	0.37500	0.37496
7	2	0.50008	0.37491	0.37491
7	3	0.63215	0.92502	0.82824
8	1	0.32120	0.35242	0.33305
8	2	0.29016	0.24818	0.29069
8	3	0.62142	0.59732	0.71258
9	1	0.91012	0.87492	0.87486
9	2	0.83807	0.87494	0.87494
9	3	0.75011	0.87501	0.87497
10	1	0.62501	0.62495	0.62498
11	1	0.40006	0.37497	0.37493
11	2	0.49007	0.37493	0.37489
11	3	0.59873	0.69820	0.64316
12	1	0.16010	0.12499	0.12502
12	2	0.54500	0.54727	0.58666
12	3	0.17111	0.12498	0.12501

Table 8.1: The scaled similarities of all the training pairs for all the basecases. These similarities should correspond with the goals in Table J.2.

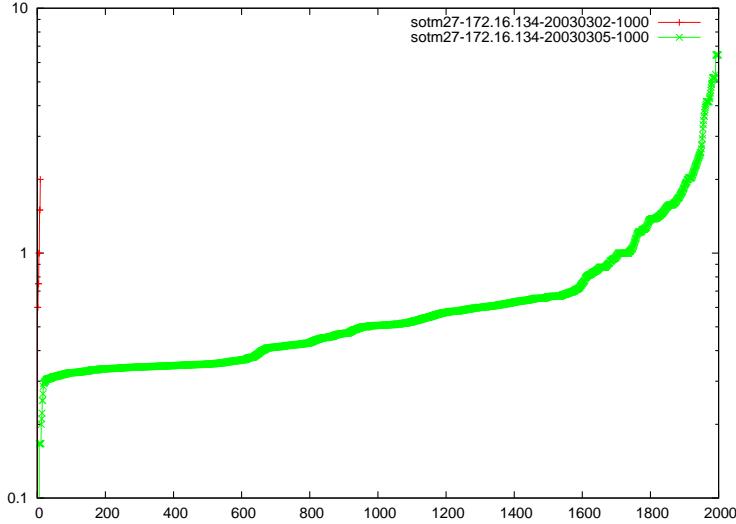


Figure 8.2: Portion of packets to privileged services in the connections over the last w seconds for basecase 4, pair 3

it more in line with the observed similarity. We are going to try changing the goal, as we wish to lower the weights of the characteristics that indicated that this pair had a high similarity, and increase the weights of the characteristics that indicated it was dissimilar.

We see a similar occurrence in basecase 7, pair 3. This time, even though the goal was much lower than it was in basecase 4, we still have a scaled similarity which is much higher across the board. When we look at Figure 8.3, we see that once again we are dealing with a trace with a lot of traffic versus one with barely any. While we knew these two would be dissimilar, we just did not anticipate how dissimilar. Most likely, it was a combination of this pair with basecase 4, pair 3, which in concert elevated the wrong weights.

Basecase 12, pair 2 is yet another occurrence. Here, even our lowest goal weight appears to be too much, given the vast disparity illustrated yet again in Figure 8.4. Basecase 8, pair 3 and basecase 11, pair 3 both appear to be more of the same – we will spare the reader yet another figure of this phenomenon. In all three cases we believe it would be prudent to set the goal down much closer to zero.

The first two pairs of basecase 8 show a very different story: in these cases there was much more traffic to compare. Figure 8.5 stands in stark contrast to the ones that came before it, as there is an appreciable amount of traffic in both traces. In fact, the characteristic distributions of the two traces seem reasonably similar, as illustrated in Figure 8.6, which can be contrasted with Figure 8.2 (keep in mind that our comparison functions scale the domain and range of the distributions). Basecase 8, pair 2 is not quite as similar, but like pair 1, much more so than the

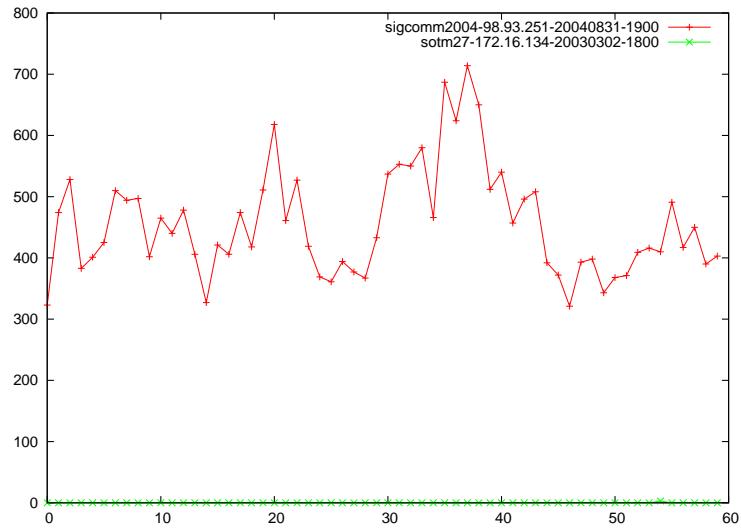


Figure 8.3: Number of connections for each second after the minute for basecase 7, pair 3

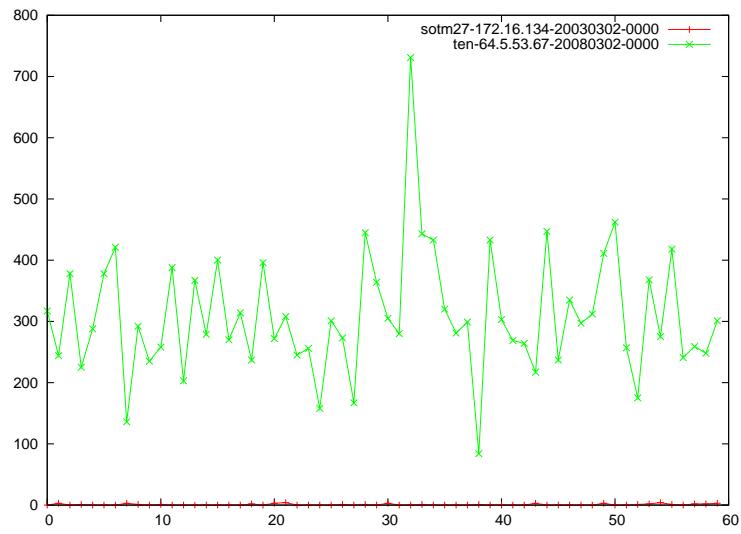


Figure 8.4: Number of connections for each second after the minute for basecase 12, pair 2

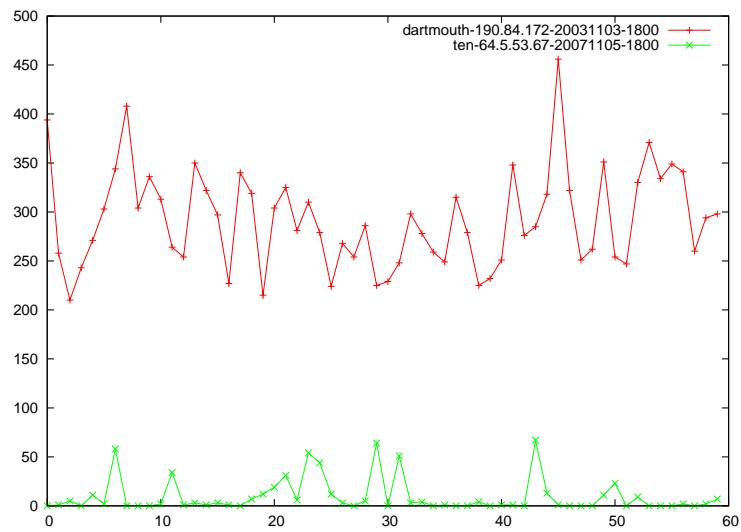


Figure 8.5: Number of connections for each second after the minute for basecase 8, pair 1

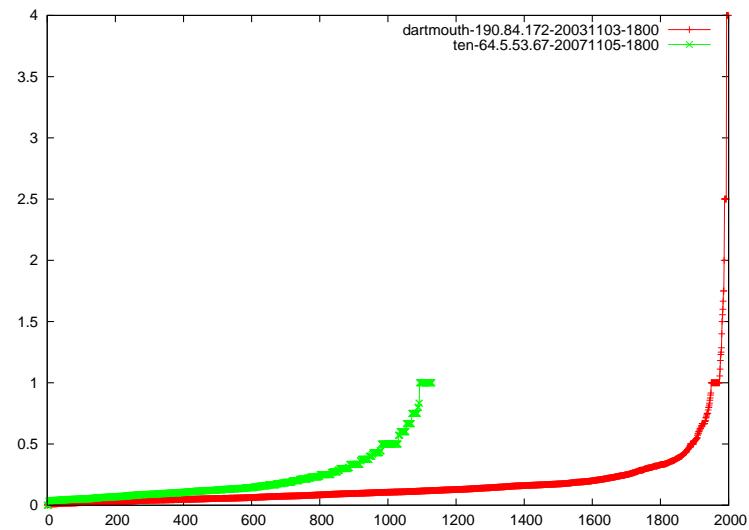


Figure 8.6: Portion of packets to privileged services in the connections over the last w seconds for basecase 8, pair 1

other pairs we have been discussing. It is probably reasonable to elevate the goals of these two pairs while keeping them under 0.25.

After making these adjustments, the problem is actually amplified, as can be seen in the first column of similarity values in Table 8.2. If we look, instead, at the second column of similarities in Table 8.2, we see that these training pairs do seem to have some value; in their absence, not only do the validated values for those pairs go even more out of whack, but some of the regular pairs deviate even more – in particular, basecase 1, pair 3 has deviated outside of the [0..1] range of similarity. Given the above results, we are going to proceed with our original set of three scores; we anticipate that the linear goals with zeros will provide the best performance on our test data, however since we have the ability to test all three sets of goals, we will do so.

8.4.3 Test cases

Obviously, the different goals resulted in different results on the test cases. We did not anticipate how much of a difference these tweaks would make. The normalized similarities along with the three scaled similarities corresponding to the three goal weights are shown in Table 8.3. In a visual analysis of these results, we are inclined to prefer the results of the linear goals without zeros, as the weights corresponding to that set of goals does not produce any negative similarities (which are outside our defined range of [0..1]). Indeed, when we actually calculate the differences between the goals and the scaled similarities, the linear goals without zeros performs best with a mean difference of 0.18755, followed by the scaled goals with a mean difference of 0.21789, and the linear goals with zeros having the worst performance with a mean difference of 0.24205.

Given this, we will move forward using the results based on the weights derived from the linear goals without zeros. Table J.3 shows the weights calculated by linear regression. Unfortunately, it only outputs the 51 largest absolute weights and the y-intercept. The assertion seems to be that the weights of the other 49 metrics are too close to zero to affect the outcome. We found a way to calculate the values for the test cases directly from the model within Weka, and compare its values with ours in Table J.4. We can not explain why such a large discrepancy exists with the two baselines which we calculate out to be zero (unless Weka is reporting a different y-intercept from what it is using). Both basecase 4, pair 2 and basecase 12, pair 2 are instances of one trace having almost no data, generating instabilities in the calculations as we saw in the section on data validation. Basecase 8, pair 1 is more puzzling, as it has a follows the same pattern as the first two training pairs of basecase 8. We will take these discrepancies into consideration as we do our analysis in the next chapter.

Base-case	pair	New Goal	New Sim	New Sim w/o near 0s
1	1	0.875	0.87499	0.87492
1	2	0.000	0.00000	0.00000
1	3	0.875	0.99702	1.01139
2	1	0.875	0.87493	0.87487
2	2	0.000	0.00000	0.00000
2	3	0.875	0.87492	0.87483
3	1	0.875	0.87500	0.87488
3	2	0.875	0.87492	0.87482
3	3	0.875	0.87493	0.87471
4	1	0.875	0.87493	0.87483
4	2	0.875	0.87501	0.87482
4	3	0.050	0.67857	0.66473
5	1	0.675	0.67496	0.67495
5	2	0.675	0.67495	0.67493
5	3	0.675	0.67490	0.67489
6	1	0.575	0.69701	0.71143
6	2	0.575	0.57499	0.57490
6	3	0.575	0.57491	0.57483
7	1	0.375	0.37491	0.37498
7	2	0.375	0.37491	0.37489
7	3	0.050	0.83228	0.78382
8	1	0.225	0.32142	0.36419
8	2	0.175	0.38263	0.53195
8	3	0.050	0.73216	0.92354
9	1	0.875	0.87493	0.87483
9	2	0.875	0.87494	0.87481
9	3	0.875	0.87495	0.87485
10	1	0.625	0.62497	0.62488
11	1	0.375	0.37490	0.37483
11	2	0.375	0.37492	0.37487
11	3	0.050	0.58376	0.60862
12	1	0.125	0.12502	0.12484
12	2	0.050	0.68601	1.01022
12	3	0.125	0.12499	0.12484

Table 8.2: The new goals and corresponding scaled similarities of all the training pairs for all the basecases. The second set of scaled similarities is produced by dropping the basecases with a goal near (but not equal to) zero from the training set; they are still included in the validation, shown here.

Basecase	Pair	Normalized Sim	Scaled Sim Scaled Goal	Scaled Sim Linear Goal	Scaled Sim Linear w/ zeros
1	1	0.69319	0.87668	0.90361	0.91042
1	2	0.61453	0.78035	0.81401	0.79634
1	3	0.68071	0.59421	0.55776	0.73861
2	1	0.60357	0.70184	0.73409	0.71750
2	2	0.00000	0.00000	0.87500	0.00000
2	3	0.68291	0.99457	0.95323	0.81368
3	1	0.71081	0.53818	0.52038	0.57072
3	2	0.55988	0.69100	0.84740	0.72695
3	3	0.59716	0.43407	0.32211	0.30577
4	1	0.72764	0.71139	0.57264	0.61129
4	2	0.40203	0.46651	0.67756	0.39465
4	3	0.00000	0.00000	0.87500	0.00000
5	1	0.60309	0.53657	0.52419	0.51724
5	2	0.50781	0.36291	0.34889	0.32604
5	3	0.00000	0.00000	0.87500	0.00000
6	1	0.73005	0.61716	0.54402	0.66293
6	2	0.60287	0.50028	0.49517	0.49201
6	3	0.59058	0.61807	0.60753	0.60309
7	1	0.57679	0.23892	0.14161	0.13491
7	2	0.55587	0.39955	0.42555	0.26887
7	3	0.61258	0.41008	0.42055	0.40197
8	1	0.39364	0.51419	0.54924	0.65569
8	2	0.49795	0.53064	0.52051	0.53249
8	3	0.37515	0.53471	0.82590	0.64334
9	1	0.64168	0.69798	0.56992	0.63654
9	2	0.69074	0.61866	0.59973	0.57320
9	3	0.70376	0.77030	0.91084	0.82976
10	1	0.60694	0.60454	0.54543	0.58089
11	1	0.38339	0.10851	0.11017	0.06443
11	2	0.60735	0.32791	0.32191	0.38154
11	3	0.51846	0.55573	0.63346	0.67264
12	1	0.42073	0.21439	0.23540	0.19132
12	2	0.41006	0.40154	0.66032	0.50716
12	3	0.45427	0.32500	0.35769	0.40713

Table 8.3: The normalized similarities with the scaled similarities for the three weights corresponding to the three different sets of goal values (all values rounded to five significant digits) for all test pairs of real data.

9

Analysis

In this chapter, we will analyze how well the above methodology fulfilled the success metrics, and hence the overall goal, laid out in chapter 3. To do so, it will provide detail on

1. The characteristics which proved useful and the degree to which each agreed with the defined difference between each pair of datasets (as determined by the weights determined on each characteristic),
2. The characteristics that did not prove useful, and an analysis of why they did not work,
3. Examples of the final reports generated for the test cases, and
4. An evaluation of the methodology's performance on the given base cases.

Additional tables and figures related to this chapter are contained in appendix M.

9.1 Useful characteristics

Table M.1 shows the weights we eventually used for our test baselines. In total, 51 of our 100 metrics are included. Of those, 28 – over half – are negative weights – including the largest one by absolute value. These negative weights are interesting because they mean that if these characteristics of the two traces had a higher normalized similarity, they were *less* likely to be alike, overall. We will begin by looking at the highest couple positive weights, then look at the highest couple negative weights.

The test cases which were closest to their goals were baseline 1, pair 2, and baseline 3, pair 2, so we will use those two pairs as examples of comparisons which worked well. The test

cases which were furthest from their goals were basecase 8, pair 3, and basecase 3, pair 3, so we will use those as examples of comparisons which did not work well.

9.1.1 Unpriv connections connection time rate

If we look at Figure K.27, we see that the distributions of the connection time rate for unprivileged connections varies greatly with many curves with an increasing slope, but some with a decreasing slope, and some that look more like stair steps. This is reassuring as the variations indicate that the distributions can actually differentiate different behaviors and traces. Figure 9.1 shows our four basecases of interest in more detail. Basecase 3, pair 2 shows an excellent example of two similar distributions. Interestingly, despite the excellent performance of basecase 1, pair 2, the similarity of its distributions is not very high, as can be seen in Figure 9.1(a). Basecase 3, pair 3 also has a fairly high similarity, as can be seen in Figure 9.1(c), which does not help to explain its poor performance. We do start to understand the poor performance on basecase 8, pair 3 by looking at Figure 9.1(d): the trace from the sotm27 dataset has very few datapoints; here we see the sotm27 trace has only zeros, matching all the zeros in the first 46% of the dsl1 trace, resulting in a 46% similarity – higher than we would expect given the environment and time differences between the traces.

9.1.2 Packet Destination IP

As with the previous characteristic, Figure K.87 shows that the distributions of Packet Destination IP addresses are highly varied, so it should provide a good measure of network similarity. Figure 9.2(a) through Figure 9.2(d) show our four basecases of interest in more detail. Both basecase 1, pair 2 and basecase 3, pair 2 show pairs of traces in which the distributions of Packet Destination IP addresses are very close, as we expect they should be. Once again, the distributions from basecase 3, pair 3 also look very similar – more so in fact than basecase 1, pair 2 and basecase 3, pair 2 – providing no insight for its poor performance. Similarly, basecase 8, pair 3 only has a hand full of data points, which results in an uncharacteristically high similarity.

9.1.3 FINs connection time rate

Figure K.32 show all the connection time rate plots for FINs. Interestingly, we see that – with only a handful of exceptions (most notably, those of the dsl1 dataset) – the shapes of the distributions within a dataset are very similar. At the same time, the shapes between different datasets can be very different, so it seems odd that this characteristic would carry a negative

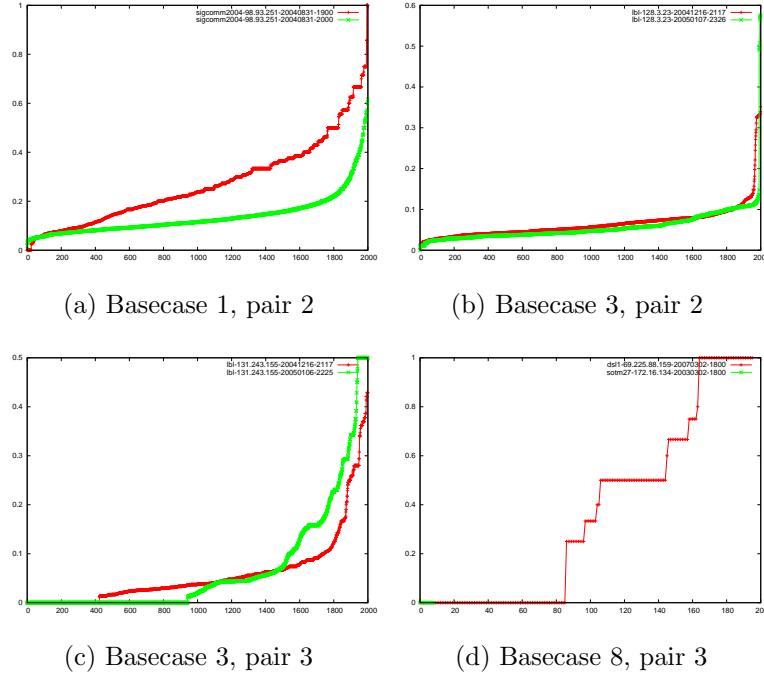


Figure 9.1: Plots of Connection time rate plot for unprivileged connections

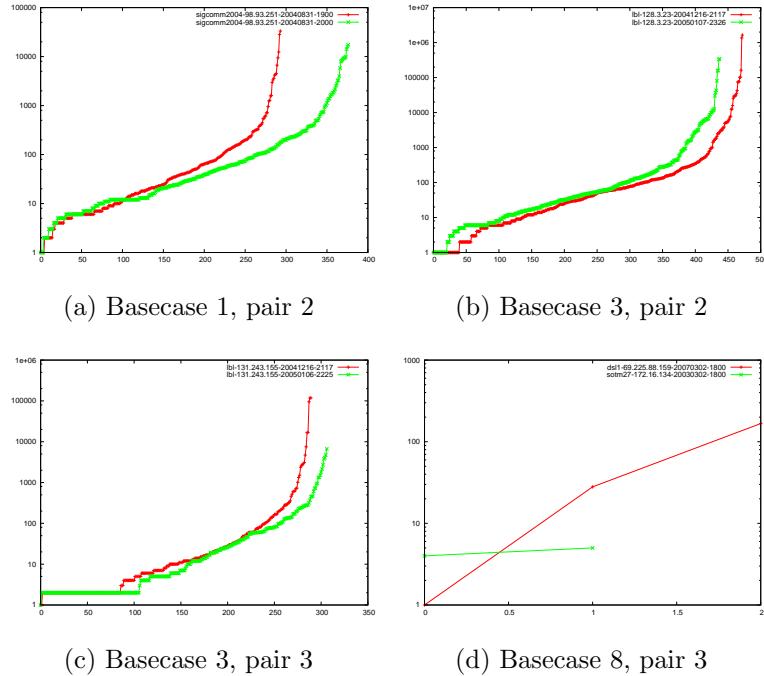


Figure 9.2: Plots of Packet Destination IP

weight. Figure 9.3(a) through Figure 9.3(d) show our four basecases of interest in more detail. The pairs of distributions in basecase 1, pair 2 and basecase 3, pair 2 show a high similarity, furthering the question of why this characteristic has a negative weight. Fortunately, the high similarity for this metric for basecase 3, pair 3 does serve to explain that pair's poor performance: the high similarity combined with the high negative weight brought brings its scaled similarity way down. In some respects, we have a similar situation with basecase 8, pair 3: due to the low number of datapoints, it has a normalized similarity of only 51% – given that it was not that high, the high negative weight does not bring down the scaled similarity as much as we would like.

9.1.4 InterPacket delta

Figure K.10 show the plots of InterPacket deltas for all traces. While not to the same extent as with FINs connection time rate, the intra-dataset distributions are typically very similar, and the inter-dataset distributions are typically very different, so once again it seems odd that it should have a large negative weight. Figure 9.4(a) through Figure 9.4(d) show our four basecases of interest in more detail. As before, basecase 1, pair 2 has a fairly high normalized similarity, which does not fit with the fact that it has a high negative weight. The normalized similarity for basecase 3, pair 2 is nowhere near as high (although that is not apparent when looking at Figure 9.4(b)), which may start to explain why it has a negative weight. Similarly, the normalized similarity for basecase 3, pair 3 is fairly low, which helps to explain why InterPacket delta has a high negative weight, but not why this pair performed so poorly. On the other hand, basecase 8, pair 3 has a very low normalized similarity, so even when combined with the high negative similarity, it helps to explain why this pair performed so poorly, but not why this metric has such a high negative weight.

9.2 Non-useful characteristics

Table 9.1 lists the metrics which are not used. When we compare these to the list of metrics which are used, we can identify a number of reasons why they were not included, such as:

1. Redundant with another metric: for instance “Packets out” is likely redundant with “Packets in”, and “Bytes service” is likely redundant with “Packets service”.
2. Better method: we see that the discrete method for “Packet TTL” was assigned a weight, whereas the orderedContinuous was not.

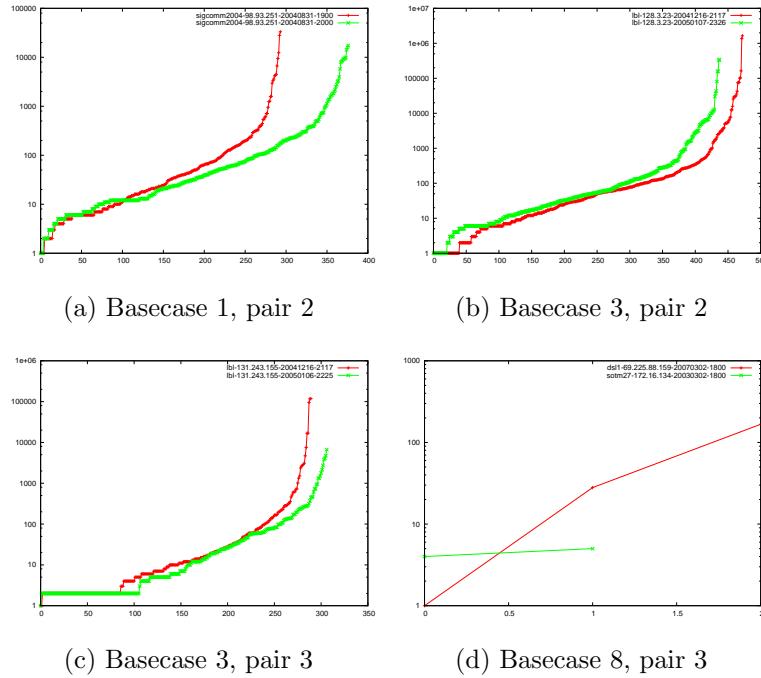


Figure 9.3: Plots of Connection time rate plot for FINs

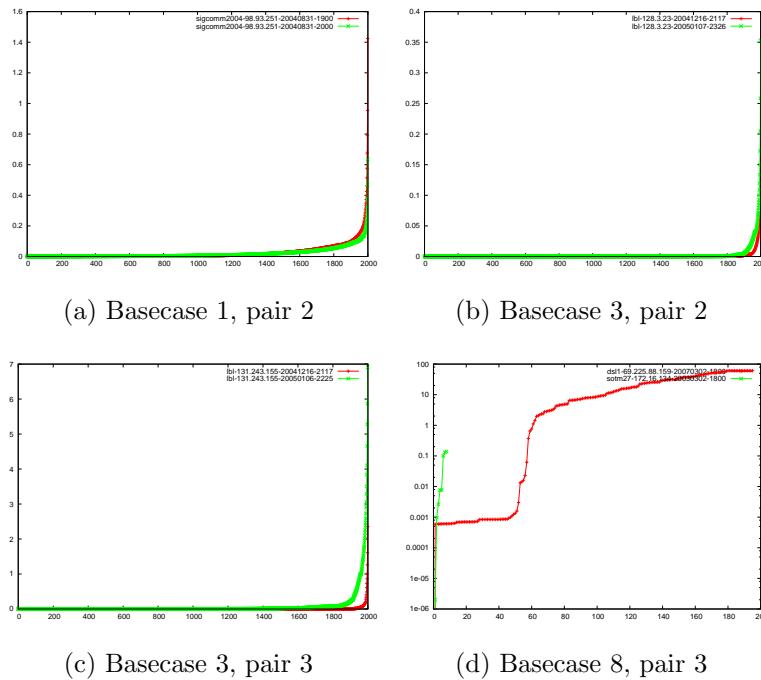


Figure 9.4: Plots of InterPacket delta

3. Complete lack of correlation: when we look at metrics such as “Packet GmHour” in Figure K.13, we see the values are all over the board. Indeed, many of the pairs are by definition of adjacent hours of traffic, meaning we will not have any similarity between these traces. Similarly, when we look at “Packet size”, they are all over the map and do not provide a correlation to how similar two traces are.

Table 9.1: Characteristics which were not assigned weights by linear regression using the scaled goals

Metric	
Packets out count	Connections in count
Bytes out count	SYN-ONLY rate ratio
Idle connection rate ratio	Bytes Service discrete
Bytes Source port discrete	Packet TTL orderedContinuous
Packet sec orderedContinuous	Packet min orderedContinuous
Packet GmHour orderedContinuous	Packet LocHour orderedContinuous
Bytes sec orderedContinuous	Bytes min orderedContinuous
Bytes GmHour orderedContinuous	Bytes LocHour orderedContinuous
Packet size orderedContinuous	Priv packets priv connection time rate sortedContinuous
SYNs connection time rate sortedContinuous	PSH connection time rate sortedContinuous
Other errors connection time rate sortedContinuous	Disconnection errors connection time rate sortedContinuous
Connection GmHour orderedContinuous	Connection LocHour orderedContinuous
Connection Priv connections rate sortedContinuous	Connection SYNs rate sortedContinuous
Connection Other errors rate sortedContinuous	Connection Disconnection errors rate sortedContinuous
Number of packets orderedContinuous	Number of packets out orderedContinuous
Number control packets rate sortedContinuous	Number bytes transferred orderedContinuous
Number bytes transferred out orderedContinuous	Number data bytes transferred out orderedContinuous
Number data bytes transferred out orderedContinuous	Fragmented packets rate sortedContinuous
Bad fragment rate sortedContinuous	Max Src Window orderedContinuous
Max Dst Window orderedContinuous	Urgent rate sortedContinuous

Continued on next page

Table 9.1 – continued from previous page

Metric	
Wrong resend rate sortedContinuous	Duplicate ACK rate sortedContinuous
Wrong ACK sortedContinuous	Wrong data packet size rate sortedContinuous
Window exceeded rate sortedContinuous	Hole rate sortedContinuous
Number reset connection orderedContinuous	Number disconnection errors orderedContinuous
Bytes Destination IP nonKeyedSortedContinuous	Packet Source IP nonKeyedSortedContinuous
Bytes Source IP nonKeyedSortedContinuous	

9.3 Reports

We have put together the `genSimReport.pl` script to generate reports of what the similarity of each pair of traces is, with details on what characteristics led to that value, how much they contributed (positive or negative), as well as the list of characteristics which were not used (for completeness sake). Appendix M contains the full text of the reports we used for the four pairs of traces used in section 9.1.

9.4 Evaluation

When we look at Table 8.3, we see some cases where the calculated value differs significantly from the goal value. To this end, we take the two sets of trace pairs which demonstrated the worst overall performance, and plot out all of their characteristics in Figure M.1 through Figure M.12 for basecase 3, pair 3, and Figure M.13 through Figure M.24 for basecase 8, pair 3.

The two pairs of traces demonstrate opposite problems in our weight assignment. Basecase 3, pair 3 is a pair of traces which is assigned a value much lower than we think it should be. One possible reason for this is that the two traces are, in fact, not as similar as we think they should be. By reviewing Figure M.1 through Figure M.12 we see a high degree of similarity in the majority of characteristics. In fact, the root of our problem seems to be a high degree of similarity in the characteristics with negative weights. There are also a number of characteristics with a high normalized similarity and a low positive or zero weight. As we noted in section 9.2, there are a number of valid reasons for this to happen, however it still leaves us questioning the validity of our weight assignment approach.

Similarly, Figure M.13 through Figure M.24 demonstrate that the two traces in basecase 8, pair 3 are quite different. Here we actually see a couple problems: besides weight assignments

which do not necessarily best reflect the similarity of the pair, there are a number of characteristics here which visibly look different between the two traces, but which have a higher normalized similarity than we would like due to the small number of data points. Intuitively, trying to characterize anything based on a small number of data points is difficult.

The common theme that we see for our poor performance is poor weight selection. In particular, the negative weights are not intuitive – in all the pairs of characteristic distributions we looked at, we did not see any indication that any characteristics were more similar if the normalized similarities were lower than others. The idea of a y-intercept – other than zero – is also rather suspect. Given these two points, we wondered if our earlier technique of determining weights independently using a linear equation was actually the right approach. A quick test showed us that this was not the case. While performance on extreme cases – such as basecase 3, pair 3 and basecase 8, pair 3 – was not as bad, overall performance (based on mean difference from goal) was lower (around 0.24 for either set of weights, whether or not empty cases were included). Similarly, we wondered if our training cases with empty traces were skewing the weight determination, but when we removed those training pairs overall performance fell.

We also considered that our problem may stem from trying to get characteristic weights represent too much – specifically differences in how networks behave over time as well as differences in the use policy of the networks themselves. We tried both splitting the training cases into those based on temporal differences from those based on network policy, however performance on the appropriate test cases was similar to most of our other tests, and still lower than our best performing weights.

As we started to split up our training set, we started to wonder if our problem might actually stem from a lack of training data. Linear regression is not typically a technique that requires a lot of training data (unlike techniques such as neural networks), however it is possible that our training set was not sufficiently representative to generate weights that would perform as well as we would like. In particular, the lower performance of basecase 8, pair 1 given that there was substantially less training data for highly dissimilar traces, lends weight to this hypothesis.

Alternatively, the problem may stem from our weight selection method. Perhaps linear regression is not the right algorithm to determine weights. Fortunately, the methodology does not dictate any particular weight selection algorithm – this detail is specific to a given instantiation of the methodology. A useful area of future investigation is if different model fitting algorithms will work better for weight determination.

10

Conclusion

This chapter summarizes our findings and presents areas for future work.

10.1 Summary of Findings

This work has presented a number of key contributions to the field of network science:

1. An extensive list of IPv4 traffic characteristics, with definitions (as found in Appendix B)
2. Methods to extract those traffic characteristics from network traces
3. Methods to model the distributions of those characteristics
4. Methods to compare the distributions of those characteristics

We will look at each of these briefly.

10.1.1 Characteristics

The complete list of IPv4 traffic characteristics is given in Appendix B. Formal definitions for these characteristics have been published in (Brugger 2007b). A basic analysis of the single value and discrete metrics appears in Chapter 5, and a basic analysis of the continuous characteristics appears in Appendix H.

10.1.2 Extracting characteristics

Following the formal definitions provided in (Brugger 2007b), we have published (Brugger 2007c) which provides a reference implementation for extracting the characteristics from network traces.

10.1.3 Modeling characteristics

We refer the reader back to Table 7.1 to see the methods used for each of the characteristics. The modeling of count, ratio, and discrete characteristics are covered in section 5.3. The modeling of sorted continuous characteristics is described in section F, ordered continuous characteristics in section G, and non-keyed sorted continuous characteristics in section H.1.7. We will review each of these definitions here.

Count characteristics

Count characteristics are modeled using a single, unsigned integer value.

Ratio characteristics

Ratio characteristics are modeled using a single real value in the range [0..1].

Discrete characteristics

Discrete characteristics are modeled as a list of key, value pairs. In practice, these pairs are stored in a hash for efficient access.

Ordered continuous characteristics

Ordered characteristics, like discrete characteristics, are modeled as a list of key, value pairs; in fact, a couple characteristics are treated as both discrete and ordered characteristics using the same model (only the comparison differs). As with discrete characteristics, these pairs are often stored in a hash for efficient access.

Sorted continuous characteristics

Sorted characteristics are modeled as a list of numbers in sorted order, from lowest to highest. Since many of the numbers may be repeated many times, we typically store it as a value, followed by the number of times it occurs in the sequence.

Non-keyed sorted continuous characteristics

Non-keyed sorted characteristics are an odd combination of ordered and sorted continuous characteristics in that – like discrete or ordered continuous characteristics – they are modeled as a list of key, value pairs (which may use a hash for efficient access), however as will be covered in the next section, they are compared like sorted continuous characteristics.

10.1.4 Comparing characteristics

As noted in the previous section, Table 7.1 shows the methods used for each characteristic, and the actual comparison of count, ratio, and discrete characteristics is described in section 5.3. The actual comparison of all the continuous characteristics is briefly covered in section 7.3. We will cover all the comparison methods here.

Count characteristics

Given two counts, x_1 and x_2 :

$$1 - \frac{|x_1 - x_2|}{x_1 + x_2}$$

Ratios characteristics

Given two ratios, x_1 and x_2 , if $(x_1 + x_2) = 0$, then the normalized similarity is defined to be one (1), otherwise it is:

$$1 - \frac{|x_1 - x_2|}{x_1 + x_2}$$

Discrete characteristics

For discrete characteristics, we first adjust the packet, connection, or byte count by $\frac{\tau_1}{\tau_2}$, where τ_1 is the count for the first trace, and τ_2 is the count for the second trace. This allows for the scale of packets, connections, or byte counts to be the same for both discrete distributions; that way we are actually comparing the distributions, not the scales of the networks, which is done with the packet, connection, and byte counts themselves.

Once that is done, we compute

$$1 - \frac{\sum_{i=1}^n \frac{|x_{(i,1)} - x_{(i,2)}|}{x_{(i,1)} + x_{(i,2)}}}{n}$$

where $|x_{(i,1)} - x_{(i,2)}|$ is the absolute value of the difference between the i^{th} measurement of the first and second traces.

Ordered characteristics

First, apply the same scaling as we do for discrete characteristics. Then, for each key, value pair, create a new set of key, value pairs equal to the original set, plus $.5 * value$ to each key adjacent to the current key, and $.25 * value$ for the two keys adjacent to those. The same comparison formula used for discrete characteristics can then be applied.

Sorted characteristics

Take the two arrays in sorted order (if an array is stored in compressed form, expand it out first), and compute

$$1 - \frac{\sum_{i=1}^n |x_{(i,1)} - x_{(j,2)}|}{n}$$

where n is the number of elements in the smaller array, m is the number of elements in the larger array, $|x_{(i,1)} - x_{(j,2)}|$ is the absolute value of the difference between the i^{th} element of the smaller array and the j^{th} element of the larger array, and $jf = i * \frac{m}{n}$ such that $x_{(j,2)} = x_{(\text{round}(jf),2)}$ if jf is within 0.00001 of that integral value, otherwise, $x_{(j,2)} = \frac{x_{(\text{floor}(j),2)} + x_{(\text{ceil}(j),2)}}{2}$.

Non-keyed sorted characteristics

As noted earlier, non-keyed sorted characteristics are modeled the same way as ordered characteristics. When we go to compare them, however, we ignore the keys in the key, value pairs by extracting all the values into an array, sorting it in ascending order, and applying the same formula as we do for sorted characteristics.

10.2 Future Work

While this work has provided a number of key contributions to the field, it has also raised a number of questions worthy of future investigation. Some of the major questions associated with overall performance of instances of the methodology follow:

1. Are there better methods to assign weights to the normalized similarities of the characteristics to find our desired scaled similarity than linear regression?
2. How much data is necessary to determine such weights?
3. What impact do second and higher order characteristics have on the results?
4. What impact does treating related UDP packets as a single connection have on the results?

5. What are the proper selections for m , n , and w (number of previous connections, number of previous packets, and number of previous seconds) when producing calculated connection attributes?

There are also a number of areas for future investigation concerning the definitions, models, and comparison methods we used for the various characteristics:

1. Should per port (or service) measurements be made on a per connection – rather than a per packet – basis?
2. Would some of the characteristics be better represented by a cumulative distribution format (CDF)?
3. Would some of the characteristics be better represented by a time series model?
4. Would modeling and comparison methods for the sorted continuous characteristics which take into account changes from one measurement to the next provide better measures of similarity?
5. Does ignoring connections which appear to already be established at the time observation begins provide better measures of similarity?
6. Does scaling the per IP characteristics (either by linear or logarithmic scaling) to match the maximum values provide better measures of similarity?
7. Does separating the per IP characteristics into internal and external IP addresses (with respect to the target network) provide better measures of similarity?

Finally, our research has produced one question which we believe will lead into further research beyond network comparison:

1. Can the models we build on characteristics for comparison purposes be used to accurately simulate a network?

We anticipate that this question is the first in a long line of questions for the application areas discussed in Chapter 2 that this research will allow the community to address.

Bibliography

- Axelsson, S. (2000, March). Intrusion detection systems: A survey and taxonomy. Technical Report 99-15, Chalmers Univ. of Technology, Göteborg, Sweden.
- Barbará, D., N. Wu, and S. Jajodia (2001). Detecting novel network intrusions using bayes estimators. In *Proc. of the First SIAM Int. Conf. on Data Mining (SDM 2001)*, Chicago. Society for Industrial and Applied Mathematics (SIAM).
- Barford, P., J. Kline, D. Plonka, and A. Ron (2002, 6–8 November). A signal analysis of network traffic anomalies. In *Proc. of the 2nd ACM SIGCOMM Workshop on Internet measurement*, Marseille, France, pp. 71–82. ACM SIGCOMM: ACM.
- Bartoletti, A. and N. A. Tang (2005, 1 April). Characterizing network services through cluster-set variations. Technical Report UCRL-TR-211020, University of California, Lawrence Livermore National Laboratory, Livermore, CA.
- Bartoletti, T. (2004, 24–27 May). Visualizations in hostile rapid scan forensics. In *DOE Computer Security Group Training Conf.*, Kansas City, MO. DOE. Available from LLNL Library as UCRL-CONF-204182.
- Brugger, S. T. (2007a). Data mining methods for network intrusion detection. *ACM Computing Surveys*. Under revision for resubmission.
- Brugger, S. T. (2007b, October). Definitions of TCP/IP connection metrics. Technical Report CSE-2007-31, University of California, Davis, Department of Computer Science, Davis, CA. <http://www.cs.ucdavis.edu/research/tech-reports/2007/CSE-2007-31.pdf>.
- Brugger, S. T. (2007c). Pcap parser and utilities. <http://sourceforge.net/projects/pcaputils>.
- Brugger, S. T. and J. Chow (2007, January). An assessment of the DARPA IDS Evaluation Dataset using Snort. Technical Report CSE-2007-1, University of California, Davis, Department of Computer Science, Davis, CA. <http://www.cs.ucdavis.edu/research/tech-reports/2007/CSE-2007-1.pdf>.

- Caswell, B. and M. Roesch (2004, 16 May). Snort: The open source network intrusion detection system. <http://www.snort.org/>.
- Chittur, A. (2001). Model generation for an intrusion detection system using genetic algorithms. High School Honors Thesis, Ossining High School. In cooperation with Columbia Univ.
- Danzig, P. B., S. Jamin, R. Cáceres, D. J. Mitzel, and D. Estrin (1992, March). An empirical workload model for driving wide-area TCP/IP network simulations. *Journal of Internetworking* 3(1), 1–26.
- Dartmouth College (2008). CRAWDAD Wireless Network Archive. <http://crawdad.cs.dartmouth.edu/data.php>.
- DatCat (2008). Internet measurement data catalog. <http://www.datcat.org/>.
- Dickerson, J. E. and J. A. Dickerson (2000, July). Fuzzy network profiling for intrusion detection. In *Proc. of NAFIPS 19th International Conference of the North American Fuzzy Information Processing Society*, Atlanta, pp. 301–306. North American Fuzzy Information Processing Society (NAFIPS).
- Erramilli, A., R. P. Singh, and P. Pruthi (1994, 6–10 June). Chaotic maps as models of packet traffic. In *Proc. 14th Int. Teletraffic Cong.*, Volume 1, North-Holland, pp. 329–338. Elsevier Science B.V.
- Ethereal, Inc. (2007). Ethereal: A Network Protocol Analyzer. <http://www.ethereal.com/>.
- Floyd, S. and E. Kohler (2003, January). Internet research needs better models. *Computer Communication Review*.
- Floyd, S. and V. Paxson (2001, August). Difficulties in simulating the Internet. *IEEE/ACM Trans. Networking* 9, 392–403.
- Gudkov, V. and J. E. Johnson (2001, 5 October). Network as a complex system: Information flow analysis. In *arXiv:nlin.CD, e-print archive on Chaotic Dynamics in Nonlinear Science*. arXiv.
- Hong, S.-S., F. Wong, S. F. Wu, B. Lilja, T. Y. Yohansson, H. Johnson, and A. Nelsson (2005, 7–8 July). Tceptransform: Property-oriented tcp traffic transformation. In K. Julisch and C. Kruegel (Eds.), *Proceedings of the Second International Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA 2005)*, Volume 3548 of *Lecture Notes in Computer Science (LNCS)*, Vienna, pp. 222–240. Springer Berlin / Heidelberg.

- Hong, S.-S. and S. F. Wu (2005, 7-9 September). On interactive internet traffic replay. In A. Valdes and D. Zamboni (Eds.), *Proceedings of the 8th International Symposium on Recent Advances in Intrusion Detection (RAID 2005)*, Volume 3858 of *Lecture Notes in Computer Science (LNCS)*, Seattle, pp. 247–264. Springer Berlin / Heidelberg.
- Javitz, H. S. and A. Valdes (1991, 20–22 May). The SRI IDES Statistical Anomaly Detector. In *Proc. 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, Oakland, CA. IEEE Computer Society.
- Kohler, E. (2004, 26 January). Ip summary dump. <http://www.icir.org/kohler/ipsumdump/>.
- Lakhina, A., K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft (2004, 12-16 June). Structural analysis of network traffic flows. In *Proceedings of the joint international conference on Measurement and modeling of computer systems*, Volume 32 of *ACM SIGMETRICS Performance Evaluation Review*, New York, pp. 61–72. ACM Special Interest Group on Performance Evaluation (SIGMETRICS): ACM.
- Lan, K.-C. and J. Heidemann (2002, July). Rapid model parameterization from traffic measurements. *ACM Trans. on Modeling and Computer Simulation* 12(3), 201–229.
- Ledesma, S. and D. Liu (2000, 21–25 August). A fast method for generating self-similar network traffic. In *Int. Conf. on Communication Technology (WCC-ICCT) Proc.*, Volume 1, Beijing, pp. 54–61.
- Lee, W. (1999). *A Data Mining Framework for Constructing Features and Models for Intrusion Detection Systems*. Ph. D. thesis, Columbia Univ.
- Lee, W., R. A. Nimbalkar, K. K. Yee, S. B. Patil, P. H. Desai, T. T. Tran, and S. J. Stolfo (2000). A data mining and CIDF based approach for detecting novel and distributed intrusions. In H. Debar, L. Mé, and S. F. Wu (Eds.), *Proc. of Third International Workshop on Recent Advances in Intrusion Detection (RAID 2000)*, Volume 1907 of *Lecture Notes in Computer Science*, Toulouse, France, pp. 49–?? Springer.
- Lee, W. and S. J. Stolfo (1998). Data mining approaches for intrusion detection. In *Proc. of the 7th USENIX Security Symp.*, San Antonio, TX. USENIX.
- Lee, W. and S. J. Stolfo (2000). A framework for constructing features and models for intrusion detection systems. *Information and System Security* 3(4), 227–261.
- Lee, W., S. J. Stolfo, and K. W. Mok (1999a, 9–12 May). A data mining framework for building intrusion detection models. In *Proc. of the 1999 IEEE Symp. on Security and Privacy*, Oakland, CA, pp. 120–132. IEEE Computer Society Press.

- Lee, W., S. J. Stolfo, and K. W. Mok (1999b, 15–18 August). Mining in a data-flow environment: Experience in network intrusion detection. In S. Chaudhuri and D. Madigan (Eds.), *Proc. of the Fifth International Conference on Knowledge Discovery and Data Mining (KDD-99)*, San Diego, CA, pp. 114–124. ACM.
- Lee, W., S. J. Stolfo, and K. W. Mok (2000). Adaptive intrusion detection: A data mining approach. *Artificial Intelligence Review* 14(6), 533–567.
- Leland, W. E., M. S. Taqqu, W. Willinger, and D. V. Wilson (1994, February). On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Trans. Networking* 2, 1–15.
- Mahoney, M. V. and P. K. Chan (2001, 10 November). PHAD: Packet Header Anomaly Detection for indentifying hostile network traffic. Technical Report CS-2001-04, Florida Institute of Technology, Melbourne, FL.
- Mahoney, M. V. and P. K. Chan (2003, 8–10 September). An analysis of the 1999 DARPA/Lincoln Laboratory Evaluation Data for network anomaly detection. In G. Vigna, E. Jonsson, and C. Krügel (Eds.), *Proc. 6th Intl. Symp. on Recent Advances in Intrusion Detection (RAID 2003)*, Volume 2820 of *Lecture Notes in Computer Science*, Pittsburgh, PA, pp. 220–237. Springer.
- McHugh, J. (2000). Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory. *ACM Trans. Information System Security* 3(4), 262–294.
- Medina, A., M. Allman, and S. Floyd (2005, April). Measuring the evolution of transport protocols in the internet. *Computer Communication Review*, 37–52.
- Mellia, M., A. Carpani, and R. L. Cigno (2002, November). Measuring IP and TCP behavior on edge nodes. In *Proc. IEEE Globecom 2002*, Taipei. IEEE.
- Miller, T. (2000, 30 August). Intrusion detection level analysis of Nmap and Queso. Technical Report 1225, SecurityFocus.
- Park, K. (1997, 7–10 December). On the effect and control of self-similar network traffic: A simulation perspective. In *Proc. of the 1997 Winter Simulation Conference*, pp. 989–996.
- Parno, B. and T. Bartoletti (2004, 9–13 August). Internet ballistics: Retrieving forensic data from network scans. Poster Presented at the 13th USENIX Security Symp., San Diego, CA.
- Paxson, V. and S. Floyd (1995, June). Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Trans. Networking* 3(3), 226–244.

- Portnoy, L., E. Eskin, and S. J. Stolfo (2001, 5–8 November). Intrusion detection with unlabeled data using clustering. In *Proc. of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001)*, Philadelphia. ACM.
- Postal, J. (1981, September). RFC 793: Transmission control protocol.
- Princeton University Cognitive Science Laboratory (2006). WordNet 3.0. <http://wordnet.princeton.edu/>.
- Ptacek, T. H. and T. N. Newsham (1998, January). Insertion, evasion and denial of service: Eluding network intrusion detection. Technical report, Secure Networks, Inc.
- Racz, P. I., T. Matsuda, and M. Yamamoto (2003, 28–30 August). Contribution of the application, transport and network layers to the self-similarity of Internet traffic. In *Proc. of 2003 IEEE Pacific Rim Conf. on Communications, Computers and Signal Processing*, Volume 2, pp. 760–763. IEEE.
- Ringberg, H., A. Soule, J. Rexford, and C. Diot (2007, 12-16 June). Sensitivity of PCA for traffic anomaly detection. In *Proceedings of the 35th joint international conference on Measurement and modeling of computer systems (SIGMETRICS '07)*, Volume 35 of *ACM SIGMETRICS Performance Evaluation Review*, San Diego, pp. 109–120. ACM Special Interest Group on Performance Evaluation (SIGMETRICS): ACM.
- Shmoo (2007). DEFCON Capture The Flag data. <http://cctf.shmoo.com/data/>.
- Sikdar, B. and K. S. Vastola (2001, 21–23 March). The effect of TCP on the self-similarity of network traffic. In *Proc. of the 35th Conf. on Information Sciences and Systems*, Baltimore, MD.
- Singh, S. and S. Kandula (2001, May). Argus - a distributed network-intrusion detection system. Undergraduate Thesis, Indian Institute of Technology.
- TCPDUMP Project (2006). TCPDUMP public repository. <http://www.tcpdump.org/>.
- University of Waikato (2007). WEKA software. <http://www.cs.waikato.ac.nz/ml/weka/>.
- Willinger, W., R. Govindan, S. Jamin, V. Paxson, and S. Shenker (2002, 19 February). Scaling phenomena in the internet: Critically examining criticality. *Proceedings of the National Academy of Sciences* 99(Suppl 1), 2573–2580.
- Willinger, W., V. Paxson, and M. S. Taqqu (1998). *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Chapter Self-similarity and Heavy Tails: Structural Modeling of Network Traffic. Boston: Birkhäuser.

Yook, S.-H., H. Jeong, and A.-L. Barabási (2002, 15 October). Modeling the Internet's large-scale topology. *Proc. of the Nat'l Academy of Sciences* 99(21), 13382–13386.

Appendix A

Similarity Calculation Code

The following is the python code used for the basic example application of the methodology in section 3.2.

```
#!/usr/bin/env python
from math import floor, ceil

def single(x1, x2):
    if (x1+x2) == 0: return 1.0
    return 1 - ((abs(x1-x2)*1.0)/(x1+x2))

def discrete(hash1, hash2):
    sum = 0
    count = 0
    while len(hash1) > 0:
        key = hash1.keys()[0]
        count += 1
        if key in hash2:
            sum += single(hash1[key], hash2[key])
            del hash2[key]
        del hash1[key]
    count += len(hash2)
    return sum/count

def updatehash(hash, origidx, newhash, newidx, factor, maxkey):
    if 0 <= newidx <= maxkey:
        if newidx in newhash:
            newhash[newidx] += factor * hash[origidx]
        else:
            newhash[newidx] = factor * hash[origidx]

def weighhash(hash):
    newhash = hash.copy()
    maxkey = max(hash.keys())
    for key in hash.keys():
        updatehash(hash, key, newhash, key-2, 0.25, maxkey)
        updatehash(hash, key, newhash, key-1, 0.5, maxkey)
        updatehash(hash, key, newhash, key+1, 0.5, maxkey)
        updatehash(hash, key, newhash, key+2, 0.25, maxkey)
    return newhash
```

```

def ordered(hash1, hash2):
    newhash1 = weighhash(hash1)
    newhash2 = weighhash(hash2)
    return discrete(newhash1, newhash2)

def sorted(arr1, arr2):
    if len(arr1) > len(arr2):
        lgarr = arr1[:]
        smarr = arr2[:]
    else:
        lgarr = arr2[:]
        smarr = arr1[:]
    lgarr.sort()
    smarr.sort()
    totalsim = 0
    count = 0
    for curidx in range(0, len(smarr)):
        curele = smarr[curidx]
        smpos = 0
        if len(smarr) > 1: smpos = curidx*1.0/(len(smarr)-1)
        lgelereal = smpos*(len(lgarr)-1)
        lgpos = int(round(lgelereal))
        if lgpos-.00001 < lgelereal < lgpos+.00001:
            lgele = lgarr[lgpos]
        else:
            lowlgele = int(floor(lgelereal))
            hilgele = int(ceil(lgelereal))
            lgele = (lgarr[lowlgele] + lgarr[hilgele]) /
                    2.0
        sim = single(curele, lgele)
        totalsim += sim
        count += 1
    return totalsim*1.0 / count

chars = ['packets', 'connections', 'bytes', 'bytes_per_service', 'pkt_'
         'minute', 'SYN_flags_conn_rate', 'ave_duration', 'control_pkt_rate',
         'packets_per_IP', 'bytes_per_IP']

packets11 = 1508
packets21 = packets11
packets12 = 1412
packets22 = 126
val1 = single(packets11, packets12)
val2 = single(packets21, packets22)
print "packets", val1, val2
normsim1 = []
normsim1.append(val1)
normsim2 = []
normsim2.append(val2)

connections11 = 17
connections21 = connections11
connections12 = 15
connections22 = 2
val1 = single(connections11, connections12)
val2 = single(connections21, connections22)
print "connections", val1, val2
normsim1.append(val1)
normsim2.append(val2)

bytes11 = 458196
bytes21 = bytes11

```

```

bytes12 = 419936
bytes22 = 10749
val1 = single(bytes11, bytes12)
val2 = single(bytes21, bytes22)
print "bytes", val1, val2
normsim1.append(val1)
normsim2.append(val2)

bps11 = {"tcp/25": 11593, "tcp/80": 385629, "tcp/993": 18154,
         "udp/53": 42820}
bps21 = bps11.copy()
bps12 = {"tcp/25": 10520, "tcp/80": 387010, "tcp/993": 18980,
         "udp/53": 41686}
bps22 = {"tcp/25": 441486, "udp/53": 16710}
val1 = discrete(bps11, bps12)
val2 = discrete(bps21, bps22)
print "bytes_per_service", val1, val2
normsim1.append(val1)
normsim2.append(val2)

pm11 = {0:398, 15:428, 30:407, 45:275}
pm21 = pm11.copy()
pm12 = {0:440, 16:398, 30:323, 44:347}
pm22 = {0:646, 30:862}
val1 = ordered(pm11, pm12)
val2 = ordered(pm21, pm22)
print "pkt_minute", val1, val2
normsim1.append(val1)
normsim2.append(val2)

sfcr11 = [0]*298 + [1]*1028 + [2]*159 + [3]*23
sfcr21 = sfcr11[:]
sfcr12 = [0]*290 + [1]*1004 + [2]*201 + [3]*13
sfcr22 = [0]*144 + [1]*1364
val1 = sorted(sfcr11, sfcr12)
val2 = sorted(sfcr21, sfcr22)
print "SYN_flags_conn_rate", val1, val2
normsim1.append(val1)
normsim2.append(val2)

ad11 = [0]*1 + [2]*10 + [5]*5 + [7]*1
ad21 = ad11[:]
ad12 = [0]*1 + [2]*11 + [5]*3 + [6]*1
ad22 = [1]*8 + [2]*8
val1 = sorted(ad11, ad12)
val2 = sorted(ad21, ad22)
print "ave_duration", val1, val2
normsim1.append(val1)
normsim2.append(val2)

cpr11 = [.25]*7 + [.45]*6 + [.71]*1 + [1]*3
cpr21 = cpr11[:]
cpr12 = [.26]*5 + [.48]*8 + [.72]*1 + [1]*3
cpr22 = [0]*8 + [.52]*8
val1 = sorted(cpr11, cpr12)
val2 = sorted(cpr21, cpr22)
print "control_pkt_rate", val1, val2
normsim1.append(val1)
normsim2.append(val2)

```

```

ppip11 = [1, 17, 291, 1199]
ppip21 = ppip11 [:]
ppip12 = [2, 22, 297, 1187]
ppip22 = [12, 180, 1317]
val1 = sorted(ppip11, ppip12)
val2 = sorted(ppip21, ppip22)
print "packets_per_IP", val1, val2
normsim1.append(val1)
normsim2.append(val2)

bpip11 = [129, 5392, 80935, 371740]
bpip21 = bpip11 [:]
bpip12 = [11, 11, 11, 458159]
bpip22 = [128, 5371, 80948, 371748]
val1 = sorted(bpip11, bpip12)
val2 = sorted(bpip21, bpip22)
print "bytes_per_IP", val1, val2
normsim1.append(val1)
normsim2.append(val2)

wf1 = [.9 / var for var in normsim1]
wf2 = [.25 / var for var in normsim2]
awf = [(wf1[lcv] + wf2[lcv]) / 2 for lcv in range(len(wf1))]
scalesim1 = [awf[lcv] * normsim1[lcv] for lcv in range(len(awf))]
scalesim2 = [awf[lcv] * normsim2[lcv] for lcv in range(len(awf))]

overall1 = sum(scalesim1) / sum(awf)
overall2 = sum(scalesim2) / sum(awf)

for lcv in range(len(chars)):
    print "%s %.4f %.4f %.4f %.4f %.4f\n" % (chars[lcv], wf1[lcv], wf2[lcv], awf[lcv],
                                                 scalesim1[lcv], scalesim2[lcv])
print "Overall_sim: %.4f %.4f\n" % (overall1, overall2)

del scalesim1[9]
del scalesim2[9]
del normsim1[9]
del normsim2[9]
del awf[9]
del chars[9]
overall1 = sum(scalesim1) / sum(awf)
overall2 = sum(scalesim2) / sum(awf)
print
print "New_Overall_sim: %.4f %.4f" % (overall1, overall2)

def mapsorted(ms1, ms2):
    sum = 0
    count = 0
    while len(ms1) > 0:
        key = ms1.keys()[0]
        count += 1
        if key in ms2:
            sum += sorted(ms1[key], ms2[key])
            del ms2[key]
        del ms1[key]
    count += len(ms2)
    return sum / count

dpc11 = {0:[2,2,5], 15:[2,2,2,5,5], 30:[0,2,5,7], 45:[2,2,2,2,5]}

```

```

dpc21 = dpc11.copy()
dpc12 = {0:[2,2,5], 15:[2,2,2,5], 30:[0,2,6], 45:[2,2,2,2,5]}
dpc22 = {15:[1], 30:[2]}
val1 = mapsorted(dpc11, dpc12)
val2 = mapsorted(dpc21, dpc22)
print "duration_per_second", val1, val2
normsim1.append(val1)
normsim2.append(val2)

dpdp11 = {"tcp/25": [2,5], "tcp/80": [0,2,2,2,5,5,5],
           "tcp/993": [2,2,5], "upd/53": [2,2,2,2]}
dpdp21 = dpdp11.copy()
dpdp12 = {"tcp/25": [2,5], "tcp/80": [0,2,2,2,5,5,6], "tcp/993": [2,2],
           "udp/53": [2,2,2,2]}
dpdp22 = {"tcp/25": [2], "udp/53": [1]}
val1 = mapsorted(dpdp11, dpdp12)
val2 = mapsorted(dpdp21, dpdp22)
print "duration_per_dest_port", val1, val2
normsim1.append(val1)
normsim2.append(val2)

wf1 = [.9 / var for var in normsim1]
wf2 = [.25 / var for var in normsim2]
awf = [(wf1[lcv] + wf2[lcv]) / 2 for lcv in range(len(wf1))]
scalesim1 = [awf[lcv] * normsim1[lcv] for lcv in range(len(awf))]
scalesim2 = [awf[lcv] * normsim2[lcv] for lcv in range(len(awf))]

overall1 = sum(scalesim1) / sum(awf)
overall2 = sum(scalesim2) / sum(awf)

chars += ["duration_per_second"]
chars += ["duration_per_dest_port"]
for lcv in range(len(chars)):
    print "%s & %4f & %4f & %4f & %4f & %4f \\\\" \
          '\\\\hline" % (chars[lcv], wf1[lcv], wf2[lcv], awf[lcv],
                      scalesim1[lcv], scalesim2[lcv])
print "-----Overall_sim & -----& -----%4f & %4f \\\\" \
      "\\hline" % (overall1, overall2)

```

Appendix B

Characteristics for comparing IPv4 networks

This list is the enumeration of possible characteristics for comparing IPv4 networks. It was produced as the first step in applying our methodology to IPv4 network trace comparison. At this stage, we are only interested in first-order characteristics. We will look at higher-order characteristics such as “frag packet rate per source IP” later.

B.1 Single value metrics

1. The total number of incoming packets
2. The total number of outgoing packets
3. The total number of incoming connections
4. The total number of outgoing connections
5. The total number of incoming bytes
6. The total number of outgoing bytes
7. SYN-ONLY rate (TCP only)
8. SYN-ACK rate (TCP only)
9. Idle connection rate (TCP only)

10. Half-open connection rate (TCP only)

B.2 Discrete metrics

11. Number of packets per protocol/dest port (or type/code for ICMP) (essential)
12. Number of packets per protocol/source port (for TCP/UDP connections)
13. Number of connections per protocol/dest port (or type/code for ICMP) (essential)
14. Number of connections per protocol/source port (for TCP/UDP connections)
15. Number of bytes transferred per protocol/dest port (or type/code for ICMP) (essential)
16. Number of bytes transferred per protocol/source port (for TCP/UDP connections)
17. TTL ¹

B.3 Continuous metrics

B.3.1 Per packet

18. Packet interarrival time
19. Number of packets that arrived at a given second after the minute (from the timestamp)
20. Number of packets that arrived at a given minute after the hour (from the timestamp)
21. Number of packets that arrived at a given hour after midnight UTC (from the timestamp)
(essential)
22. Number of packets that arrived at a given hour after midnight Local (from the timestamp)
(essential)
23. Number of packets that arrived at a given day of the week (from the timestamp) (essential)
24. Number of bytes that arrived at a given second after the minute (from the timestamp)
25. Number of bytes that arrived at a given minute after the hour (from the timestamp)

¹We also model this as a continuous metric, which we think it will work better as, but it is also possible to model it as a discrete metric, so we will try it and see how well it works.

26. Number of bytes that arrived at a given hour after midnight UTC (from the timestamp) (essential)
27. Number of bytes that arrived at a given hour after midnight Local (from the timestamp) (essential)
28. Number of bytes that arrived at a given day of the week (from the timestamp) (essential)
29. Packet sizes
30. TTL

B.3.2 Per packet, time rates

31. Number of packets over the past w seconds
32. Number of packets to privileged services versus the number of packets over the past w seconds
33. Number of packets to unprivileged services versus the number of packets over the past w seconds
34. Number of connections over the past w seconds
35. Number of connections to privileged services versus the number of connections over the past w seconds
36. Number of connections to unprivileged services versus the number of connections over the past w seconds
37. Number of privileged connections versus the number of packets to privileged services over the past w seconds
38. Number of unprivileged connections versus the number of packets to unprivileged services over the past w seconds
39. Number of SYN flags versus the number of connections over the past w seconds
40. Number of RST flags versus the number of connections over the past w seconds
41. Number of FIN flags versus the number of connections over the past w seconds
42. Number of connections versus the number of PSH flags over the past w seconds

- 43. Number of establishment errors versus the number of connections over the past w seconds
- 44. Number of other errors connections versus the number of connections over the past w seconds
- 45. Number of disconnection errors connections versus the number of connections over the past w seconds
- 46. Average duration over the past w seconds

B.3.3 Per packet, packet rates

- 47. Percentage of packets to privileged services out of the past n packets
- 48. Percentage of packets to unprivileged services out of the past n packets

B.3.4 Per connection initiation

- 49. Connection interarrival time
- 50. Number of connections initiated at a given second after the minute (from the timestamp)
- 51. Number of connections initiated at a given minute after the hour (from the timestamp)
- 52. Number of connections initiated at a given hour after midnight UTC (from the timestamp)
(essential)
- 53. Number of connections initiated at a given hour after midnight Local (from the timestamp)
(essential)
- 54. Number of connections initiated at a given day of the week (from the timestamp) (essential)
- 55. Source port ² (TCP/UDP only)

²We also look at the number of connections per source port as a discrete metric, which we think will be more useful; we also want to try modeling source ports as a continuous metric based on the fact that most OSes assign them sequentially.

B.3.5 Per connection initiation, connection rates

56. Number of connections versus the number of packets over the past n connections
57. Number of connections to privileged services versus the number of connections over the past n connections
58. Number of connections to unprivileged services versus the number of connections over the past n connections
59. Number of privileged connections versus the number of packets to privileged services over the past n connections
60. Number of unprivileged connections versus the number of packets to unprivileged services over the past n connections
61. Number of SYN flags versus the number of connections over the past n connections
62. Number of RST flags versus the number of connections over the past n connections
63. Number of FIN flags versus the number of connections over the past n connections
64. Number of connections versus the number of PSH flags over the past n connections
65. Number of establishment errors versus the number of connections over the past n connections
66. Number of other errors versus the number of connections over the past n connections
67. Number of disconnection errors versus the number of connections over the past n connections
68. Average duration over the past n connections

B.3.6 Per connection close

69. Number of packets
70. Number of packets sent
71. Number of packets received
72. Connection duration

73. Control packet rate
74. Data packet rate
75. Number of bytes
76. Number of bytes sent
77. Number of bytes received
78. Number of data bytes
79. Number of data bytes sent
80. Number of data bytes received
81. Frag packet rate
82. Wrong frag packet rate
83. Max source window size (TCP only)
84. Max destination window size (TCP only)
85. Urgent packet rate (TCP only)
86. Resend packet rate (TCP only)
87. Wrong resend packet rate (TCP only)
88. Duplicate ACK packet rate (TCP only)
89. Wrong ACK (TCP only)
90. Wrong data packet size rate (TCP only)
91. Window exceeded (TCP only)
92. Hole rate (TCP only)
93. Connection errors (TCP only)
94. Reset connection (TCP only)
95. Other errors (TCP only)
96. Disconnection errors (TCP only)

B.3.7 Per IP

97. Number of packets per source IP
98. Number of packets per destination IP
99. Number of connections per source IP
100. Number of connections per destination IP
101. Number of bytes transferred per source IP
102. Number of bytes transferred per destination IP

We will investigate both using and not using virtual UDP connections, where the UDP packets between the same hosts for the same service with a maximum interpacket latency below some timeout threshold (such as 150 seconds) is treated as a single logical connection.

Appendix C

Initial test data

This appendix describes how the first five basecases, described in chapter 4 are built in detail.

C.1 Base case 1

We started with one hour of data from a Linode owned and controlled by the author, and only used by him. The hour of data spanned from 08:00-09:00 PDT on a normal Friday morning. It consists of 1818 packets of data to 30 other hosts. It was anonymized such that the fourth byte of the IP addresses on the same network as the Linode host were randomized, except for those IP addresses ending in “.0”, “.1”, or “.255”, which have special meaning in network administration¹. Of course, given that in this case the data was collected on a single host, and not the entire network, it is easy to identify which host in the anonymized data was the Linode where the collection was performed; however, without additional data, one can not tell what the real IP address of this host is. The anonymization was performed by the `simpleAnon` program, written by the author, although there are other scripts and programs which could have been used to the same effect.

C.2 Base case 2

The second base case used the same hour of data described above. Using the `modTime` program, written by the author, the timestamps in the tcpdump file were changed to appear 1

¹“.0” denotes the network number, although it is occasionally used as a gateway, “.1” is the typical address for the network gateway, and “.255” is the broadcast address

hour (3600 seconds) after they actually occurred, resulting in a tcpdump file which appeared to span from 09:00-10:00 PDT on the same Friday.

C.3 Base case 3

Base case three was trickier to create the data for. We started with the same hour of tcpdump data. We then performed the following steps:

1. Using the `splitHostStreams` program, written by the author, the tcpdump data was split into 25 separate tcpdump files, one for each host the Linode communicated with.
2. An attempt to connect to NetBIOS on the Linode², consisting of two packets from 64.2.90.106 was removed.
3. A mail connection³ from 24.177.236.232, consisting of 23 packets, was removed.
4. We took one hour of data from exactly one week later (08:00-09:00 PDT on the following Friday), and modified it to look like it was collected at the same time as the original hour of data.
5. Using the `splitHostStreams` program again, we split the modified data from one week later into separate files for the 24 hosts which the Linode communicated with at that time.
6. An attempt to connect to UDP port 1026⁴, and another attempt to connect to NetBIOS, both from 63.5.53.67, were included in our data set.
7. A mail connection from 218.51.223.195, consisting of 24 packets, was included in our data set.
8. Three of the connections, consisting of one attempt to access NetBIOS, and two SMTP connections, with 49 packets altogether, had the IP addresses of the remote hosts modified to a random value using the `modHost` program, written by the author.

²The Linode does not run the NetBIOS service, so this connection was almost certainly made with malicious intent.

³The Linode does run the SMTP service, so this was a legitimate connection from the session layer perspective, although we do not have the data portion of that connection, so we can not make any judgments as to whether this was a legitimate connection from the application layer perspective (in other words, it may have been spam).

⁴Port 1026 is registered for the “calender access protocol” (sic), however we expect that the remote host expected to find a client that had used the first available non-privileged port.

9. We jittered the time on most of the connections by adjusting the connections with four of the remote hosts forward one second, adjusting four forward two seconds, two forward three seconds, four backwards one second, four backwards two seconds, and two backwards two seconds.
10. Using `ipsumdump` version 1.59(Kohler 2004), we collated the dumps from the separate hosts back into a single `tcpdump` file consisting of 1819 packets between the Linode and 25 remote hosts.
11. Since `ipsumdump` does not sort the times within a given stream, we used the `fixTime` program, written by the author, to sort the packets by time, down to the microsecond.
12. Finally, using the `modTime` program, all the timestamps in the dump file were moved forward one hour.

C.4 Base case 4

Once the procedure was established for base case 3, this base case was easier to build. This time we started with the host streams prior to collation from base case 3.

1. The 526 SSH packets between the Linode and 128.115.223.191 were replaced with the 547 SSH packets from one week later.
2. Two packets from 202.97.181.106 which originate from port 80 and are sent to a high numbered port were removed. These packets were most likely network probes, as they are not part of any HTTP connection to the Linode, the lack of any SYN flags indicates that they are not backscatter from any attack against that remote host, and the use of port 80 as the source port would likely allow these packets through any stateless firewall.
3. A mail connection from 66.98.57.92, consisting of 40 packets, was removed.
4. From the following week's data (modified as above), a mail connection from 66.63.182.51, consisting of 27 packets, was added.
5. A single packet from 65.35.108.106, attempting to establish a connection to port 5110 (which the Linode does not run), was added.

6. Three of the connections, consisting of two UDP packets, apparently destined for a Doom server⁵, two attempts to access NetBIOS, and 48 packets from a Tor⁶ connection, had the IP addresses of the remote hosts modified to a random value using the `modHost` program.
7. We jittered the time on most of the connections by adjusting the connections with four of the remote hosts forward two seconds, adjusting four forward four seconds, two forward six seconds, five backwards two seconds, four backwards four seconds, and two backwards six seconds. The reordering of hosts due to added, removed, and modified connections, and the use of a different order of adding or removing time, prevented the same connections from all being moved forward or backward in time.
8. Using `ipsumdump` version 1.59(Kohler 2004), we collated the dumps from the separate hosts back into a single `tcpdump` file consisting of 1826 packets between the Linode and 25 remote hosts.
9. The `fixTime` program was used to sort the packets, by time, down to the microsecond.
10. Finally, using the `modTime` program, all the timestamps in the dump file were moved forward one hour.

C.5 Base case 5

This base case was significantly harder than the previous ones, as a lot of work had to be done to minimize the similarities between the datasets.

1. We started with the data file for base case 4, as it already deviated significantly from the original data file.
2. The `fixTime` program was used to sort the packets, by time, down to the microsecond. This was primarily a housekeeping measure in the previous base cases; here it is necessary, as the `modRandTime` program expects the packets to be sorted in order to calculate jitter.
3. Ten copies of that file were created, each one processed by the `simpleAnon` program. By collating these ten copies together, we should appear to have a network with ten times the number of hosts and amount of data.

⁵Doom is a multiplayer first person shooter game. The Linode has never been used as a Doom server.

⁶Tor is the next-generation Onion router protocol used for anonymous communication.

4. All ten of these files were further processed using the `modHost` program, to make all the hosts external to the home network appear to be different.
5. The ten files were then processed using the `differentiateServices` program, written by the author, to ensure that none of the ports (source or dest) in the original data file, were used in the new data file. By processing each of the ten data files separately, each one itself has a different set of ports (services) used.
6. The ten data files were then split using the `splitHostStreams` program, with all of the component streams placed in the same directory.
7. In order to change the service mix, five of the stream files had five copies made of them, with each copy processed by `modHost` to change the external host. Another five files had four copies made, five had three copies made, five had two copies made, and five had one copy made, all using `modHost` to change the external hosts. The files that were picked to copy were selected at random by the `random_file.pl` script, so some of the copies were, themselves, copied.
8. Using the `random_file.pl` script, 75 of the host streams were removed.
9. Using the `modRandTime` program, written by the author, the times of all of the host data streams were moved randomly in time to any point between an hour before to an hour after they actually occurred. Further, jitter was introduced to the intrapacket timings, with a jitter value picked randomly from a normal distribution with a mean of zero and a variance equal to the number of micro seconds since the last packet. The jitter was applied cumulatively to the stream – that is, the new interpacket arrival time was added onto the modified time, not the real time for the previous packet. Finally, the interarrival times between packets was halved, to make the packet rate appear to be double what it was in the real data.
10. Using the `randPktSize` program, written by the author, the sizes of all the packets were changed. Any packets that were smaller than 256 bytes were made a random size larger than 256 bytes, and vice-versa. This should result in a very interesting – and unrealistic – network dynamic, where the data packets are very small, and the ACKs are very large. What will really help differentiate the real and artificial trace will probably end up being that the real trace has a maximum packet size of 1500, due to contemporary layer 2 networks, whereas the artificial trace will have many packets with sizes well above this. Additionally, if the “Don’t Fragment” flag was set on the packet, it was unset, a random maximum fragment

size was selected, and the packet was fragmented appropriately. Conversely, if the “Don’t Fragment” flag was not set, it was set. It should be noted that the only fields in the packet that were modified were the IP packet length, flags, fragment offset, and the TCP maximum segment size or the ICMP data length or UDP data length, where applicable. We did not modify the TCP window sizes, or the sequence and acknowledgment numbers, meaning that the resulting streams will be logically inconsistent, however our application of the network comparison methodology does not examine these aspects of connections, making it sufficient for our purposes.

11. With the packet sizes themselves changed, we proceeded to increase the number of packets in each connection, using the `dblConnSize` program, written by the author. This program repeats the last two packets in a connection when they represent host A sending a packet to host B, followed by host B sending a packet to host A. These repeated connections are identical to the original connections, save for the timestamps, which are set to be one and two microseconds after the original second packet, and the removal of any SYN flags.
12. Using `ipsumdump` version 1.59(Kohler 2004), we collated the dumps from the separate hosts back into a single `tcpdump` file consisting of 52856 packets or fragments between the 26 hosts in the artificial network and 227 remote hosts.
13. The collated dump file was split into three files, one for each hour that the data covered, using the `tcpTimeslice` program. Two of the files were then modified using the `modTime` program, such that all the packets appear to be captured in the same hour as the original data file. All three files were collated back together using `ipsumdump` to create the final file used for the base case.

Appendix D

Count and Discrete Metrics on Initial Base Cases

The following is our original analysis on singular and discrete metrics, including our treatment of Per IP metrics as discrete metrics (when only the observed network is anonymized), linear versus non-linear weighting, and only using count characteristics as single value metrics (no ratio metrics). References are made to the scripts developed before the incorporation of continuous characteristics necessitated more robust approaches.

Given the first five base cases, we proceeded to develop scripts to extract the characteristics, find the similarity score for each characteristic, and develop the weights for each similarity to determine the overall similarity score. The first five base cases, as covered in the previous chapter, are all constructed by hand based on one hour of real network data. By using constructed data, we know what to expect, reducing surprises during the development of the methodology.

All numbers in this chapter have been rounded to four digits after the decimal place for legibility. Actual calculations were performed using single-precision floating point numbers.

D.1 Count characteristics

Rather than try to handle all identified characteristics of network data simultaneously, we decided to start simply and look at basic count-based characteristics of the data.

Measure	Trace 1	Trace 2	Norm Sim	Scaled Sim
Packets in	922	922	1.0	0.9260
Packets out	604	604	1.0	0.9261
Connections in	39	39	1.0	0.9296
Connections out	15	15	1.0	0.925
Bytes in	846380	846380	1.0	0.9221
Bytes out	71205	71205	1.0	0.4754

Table D.1: Count characteristics and Similarity for Basecase 1. Similarity Goal was 1.0 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.

D.1.1 Measuring

There are six characteristics from each network trace which are simple counts:

1. Bytes sent into the target network
2. Bytes sent out from the target network
3. Packets sent into the target network
4. Packets sent out from the target network
5. Connections into the target network
6. Connections out from the target network

We did not examine intranet traffic (traffic where the source and destination were both on the target network).

The `buildDistros` program (written by the author) takes a pcap dump file and produces the above counts.

We did not adjust the measured counts of the second network to account for differences in network sizes, as these are the characteristics we use to differentiate two networks based on size alone, and they will be used to adjust other characteristics, below.

The six counts for both traces are shown in tables Table D.1 through Table D.5. “Trace 1” is always our original one hour of data, with “Trace 2” being the data modified, as covered in the previous chapter.

Measure	Trace 1	Trace 2	Norm Sim	Scaled Sim
Packets in	922	922	1.0	0.9260
Packets out	604	604	1.0	0.9261
Connections in	39	39	1.0	0.9296
Connections out	15	15	1.0	0.925
Bytes in	846380	846380	1.0	0.9221
Bytes out	71205	71205	1.0	0.4754

Table D.2: Count characteristics and Similarity for Basecase 2. Similarity Goal was 1.0 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.

Measure	Trace 1	Trace 2	Norm Sim	Scaled Sim
Packets in	922	924	0.9989	0.9248
Packets out	604	603	0.9991	0.9253
Connections in	39	39	1.0	0.9296
Connections out	15	15	1.0	0.925
Bytes in	846380	847473	0.9994	0.9287
Bytes out	71205	71115	0.9994	0.5003

Table D.3: Count characteristics and Similarity for Basecase 3. Similarity Goal was 0.9 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.

Measure	Trace 1	Trace 2	Norm Sim	Scaled Sim
Packets in	922	926	0.9978	0.9236
Packets out	604	608	0.9967	0.9230
Connections in	39	40	0.9873	0.9129
Connections out	15	15	1.0	0.925
Bytes in	846380	847237	0.9995	0.9273
Bytes out	71205	78601	0.9506	2.3238

Table D.4: Count characteristics and Similarity for Basecase 4. Similarity Goal was 0.8 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.

Measure	Trace 1	Trace 2	Norm Sim	Scaled Sim
Packets in	922	12969	0.1327	0.0997
Packets out	604	10368	0.1101	0.0995
Connections in	39	436	0.1642	0.0983
Connections out	15	173	0.1596	0.1000
Bytes in	846380	202292725	0.0083	0.0999
Bytes out	71205	229929690	0.0006	0.0250

Table D.5: Count characteristics and Similarity for Basecase 5. Similarity Goal was 0.1 . The “Norm Sim” column gives the normalized, calculated similarity based on the two counts, and the “Scaled Sim” column gives the similarity after being scaled using the non-linear weight given in Table D.6. Composite scaled similarity score given in Table D.7.

D.1.2 Normalized Similarity

Given the six counts for the five pairs of base case network traces, we evaluated the similarity of each characteristic using the formula

$$1 - \frac{|x_1 - x_2|}{x_1 + x_2}$$

as the normalized similarity between the two. As discussed in chapter 3, this is essentially the Jaccard coefficient. This value will be limited to the range [0, 1] (hence, “normalized” similarity), as long as all of our measurements are positive numbers (which in the domain of network measurements, they should be).

The `normCompare.pl` Perl script takes the counts from both traces in a basecase, and produces the normalized similarity for each characteristic. Tables Table D.1 through Table D.5 show these normalized similarities. As expected, the normalized similarity for the first two basecases is 1.0 for all characteristics. For the third basecase, the count characteristics of the modified trace are very close to the original trace, resulting in normalized similarities around 0.99, with no change in the number of connections in or out. For the fourth basecase, there was a slight change in the number of incoming connections, putting it also in the 0.99 range. Three of the other counts remained in the 0.99 range with a slight decrease in similarity values over basecase 3, however the count of “bytes in” actually went down over basecase 3, resulting in a slightly higher normalized similarity value. This is actually somewhat fortuitous, as we expect to see pairs of traces that are more dissimilar, even though a minority of the characteristics are more similar.

The normalized similarities for the count characteristics of basecases 3 and 4 do give us pause however, as they are so close to 1.0 . This indicates that the modified traces are actually much closer to the original trace than intended. We will return to this in the next subsection when we look at the scaled similarities.

Fortunately, the modified trace for the fifth basemode turned out extremely dissimilar from the original trace, as intended. Four of the count characteristic normalized similarities were in the range 0.10 – 0.17, and the byte counts were down below 0.01 . We expect this to be sufficiently dissimilar to be useful when looking at scaled similarities.

D.1.3 Scaled Similarity

The first five base cases were constructed such that the first two basemodes should have scaled similarities of 1.0, which matches the normalized similarities for both pairs of traces. We want the third pair of traces to have a scaled similarity of 0.9 . Given that the mean normalized similarity of the third basemode is 0.9995, we could just multiply each normalized similarity by a correction factor (the mean value of which would be about 0.9005). This becomes problematic when we consider the fourth basemode, which was designed to be twice as dissimilar as the third basemode. It would require a mean correction factor of 0.8092 . Hence, in order to archive both of our target scaled similarities, we can not use constant correction factors, however we could achieve both if we used a linear correction factor $y = a + bx$ where x is the normalized similarity and y is the scaled similarity. The y -intercept would be at a , and b would be the slope of the line. This would throw off the scaled similarities for the first two base cases, however; and one can easily see that it would be vastly off for the fifth basemode as well. We would also like to define the y -intercept of such a line to be zero, as we can be fairly certain that if a normalized similarity is zero, the scaled similarity should be as well.

If we actually plot out x and y , as above, for each characteristic, we see that the points roughly follow a parabola in the first quadrant from (0, 0) to (1, 1). We can use a least squares fit to find the quadratic $y = a + bx + cx^2$. Setting $a = 0$, we just have to solve

$$c = \frac{\sum_{i=1}^n xy - \frac{\sum_{i=1}^n x^2 \sum_{i=1}^n y}{\sum_{i=1}^n x}}{\sum_{i=1}^n x^3 - \frac{\sum_{i=1}^n x^2 \sum_{i=1}^n x^2}{\sum_{i=1}^n x}}$$

$$b = \frac{\sum_{i=1}^n y - c \sum_{i=1}^n x^2}{\sum_{i=1}^n x}$$

We compute these values using the `evalSimsNonlinear.pl` Perl program, written by the author, and get the factors shown in Table D.6 for each of the six count measures.

A cursory examination of the factors shows that the curves to fit the measurements go well outside our expected bounds (between zero and one). Sure enough, we see that if we use these factors to compute the scaled similarities (shown in tables Table D.1 through Table D.5), we get mean scaled similarity scores that are far from the mark (Table D.7). In fact, basemode 3

Measure	b	c
Packets in	0.7239	0.2021
Packets out	0.9012	0.0249
Connections in	0.5335	0.3961
Connections out	0.5700	0.3550
Bytes in	12.0834	-11.1614
Bytes out	40.3593	-39.8838

Table D.6: The factors used for the non-linear scaling of the normalized similarities to produce the scaled similarities. The scaled similarity was defined as $bx + cx^2$, where x was the normalized similarity.

Basecase	Goal Sim	Scaled Sim
Basecase 1	1.0	0.8507
Basecase 2	1.0	0.8507
Basecase 3	0.9	0.8556
Basecase 4	0.8	1.1559
Basecase 5	0.1	0.0871

Table D.7: The goal similarity and composite scaled similarities for the first five base cases. The composite scaled similarity for each basecase was determined by taking the mean of the scaled similarities for all of the measures.

actually has a higher similarity score than the first two, and basecase 4 has a score higher than the maximum (1.0, or perfectly similar).

Now, we could try other scaling functions, or different ways to find the weights, but the fact of the matter is that this is a situation we expect to see with real data frequently: the actual similarity between two traces does not match our expected similarity. So instead of trying to warp the normalized similarities to match our expectations, we will change our expectations to more closely align with reality. Table D.9 shows the new goal similarities, determined using the mean value of the normalized similarities of the six count measurements.

When we reran the `evalSimsNonlinear.pl` program with the new goals and got the

Measure	b	c
Packets in	0.7239	0.2021
Packets out	0.9012	0.0249
Connections in	0.5335	0.3961
Connections out	0.5700	0.3550
Bytes in	12.0834	-11.1614
Bytes out	40.3593	-39.8838

Table D.8: The factors used for the non-linear scaling of the normalized similarities to produce the scaled similarities, based on the new goals. The scaled similarity was defined as $bx + cx^2$, where x was the normalized similarity.

Basecase	Goal Sim	Scaled Sim
Basecase 1	1.0	0.9212
Basecase 2	1.0	0.9212
Basecase 3	0.9995	0.9261
Basecase 4	0.9887	1.2314
Basecase 5	0.0959	0.0842

Table D.9: The new goal similarity and composite scaled similarities for the first five base cases. The composite scaled similarity for each basecase was determined by taking the mean of the scaled similarities for all of the measures.

factors shown in Table D.8. Using these factors we reran the `evalCalcNonlinear.pl` program and got the composite scaled similarities shown in Table D.9. Just looking at the factors, it is apparent that changing the goal similarities did not have a huge impact. Sure enough, the new composite scaled similarities might have shifted a bit such that the 1.0 scores are actually closer to 1.0, however we see the same problem in that the scaled similarity score for basecase 3 is higher than the first two basecases, and the score for basecase 4 continues to exceed the maximum.

If we take a closer look for these odd composite scaled similarity scores, it becomes apparent that it is the “bytes out” scaled similarity score that is skewing the mean value. The cause of such erroneous scaled similarities is the inverted parabola scaling function used for the “bytes out” measure. This, in turn, was caused by the least squares fit trying to map the tiny value recorded for basecase 5 (0.0006) to a similarity of 1.0 or 0.0959.

This is even more instructive than the realization that our expected similarities may not be close to the actual similarities: it demonstrates how not all characteristics are useful for quantifying the similarities between networks. This is not to say that “bytes out” will not be a useful comparison characteristic on some networks, just that it is not useful for comparing our artificially constructed network traces. We might also give pause looking at “bytes in”, given its inverted parabola scaling function, however its parabola is much shallower, and the scaled similarities it is producing are close to our expected values.

Given this, we will re-ran the `evalCalcNonlinear.pl` program without the “bytes out” measurement, and got the values shown in Table D.10.

The removal of the “bytes out” characteristic certainly has a drastic effect: all of the composite scaled similarities are within the range of [0, 1], and are all much closer to their goals. We do note, however, that the scaled similarity for basecase 3 is still higher than basecases 1 and 2. We suspect this is due to the inverted parabola scaling function for “bytes in”, so we will remove that as well and get the values shown in Table D.11.

Finally, we have a set of scaled similarities which at least relatively correspond to our constructed basecases. We must keep in mind, however, that these composite scores are based

Basecase	Goal Sim	Scaled Sim
Basecase 1	1.0	0.9979
Basecase 2	1.0	0.9979
Basecase 3	0.9995	0.9987
Basecase 4	0.9887	0.9938
Basecase 5	0.0959	0.0959

Table D.10: The new goal similarity and composite scaled similarities for the first five base cases, without the “bytes out” characteristic. The composite scaled similarity for each basecase was determined by taking the mean of the scaled similarities for all of the measures.

Basecase	Goal Sim	Scaled Sim
Basecase 1	1.0	0.9987
Basecase 2	1.0	0.9987
Basecase 3	0.9995	0.9981
Basecase 4	0.9887	0.9924
Basecase 5	0.0959	0.0959

Table D.11: The new goal similarity and composite scaled similarities for the first five base cases, without the “bytes out” or “bytes in” characteristics. The composite scaled similarity for each basecase was determined by taking the mean of the scaled similarities for all of the measures.

on just four characteristics. This is fine as the next step is to incorporate the 11 discrete characteristics.

D.2 Discrete characteristics

Discrete characteristics in network data are those that we can keep a count per some key item. For example, the number of packets per source IP address. If the same source IP sent the same number of packets in two different traces, that portion of that characteristic has a similarity of 1.0. If a source IP sent packets to one address but not another, then the similarity is 0.0. Any percentage in between is scaled using the same formula as the count characteristics. The normalized similarity of the entire characteristic is the mean of all the individual parts.

There is one more adjustment that we make here: if one network has double the network traffic of another, it should be no surprise that it has double the number of packets from a well known Internet site. Moreover, we have already accounted for this with our “packets in” count measurement, so in order to avoid having network size skew our measurements, we multiply each count in the discrete measurement by a scaling factor inversely proportional to the size difference of the same size measurement for the second network. To extend our example of packets per source IP, if the second trace has twice the total number of packets, then we multiply the number

of packets from each source IP by 0.5 for the second trace.

D.2.1 Measuring

The discrete characteristics that we have identified are:

1. Source IP based on the number of packets
2. Source IP based on the number of connections
3. Source IP based on the number of bytes transferred
4. Destination IP based on the number of packets
5. Destination IP based on the number of connections
6. Destination IP based on the number of bytes transferred
7. Protocol/dest port (or ICMP type/code) based on the number of packets
8. Protocol/dest port (or ICMP type/code) based on the number of connections
9. Protocol/dest port (or ICMP type/code) based on the number of bytes transferred
10. TCP/UDP Source Port based on the number of packets
11. TCP/UDP Source Port based on the number of connections
12. TCP/UDP Source Port based on the number of bytes transferred
13. TTL Values

Rather than modify the existing `buildDistros` program to handle numerous maps of discrete characteristics, we put together the `decodeTcpdump` program, which decodes the relevant characteristics from each packet. We then created a `buildDistros.pl` Perl script, which did essentially the same thing as the prior `buildDistros` program, with the additional support of discrete characteristics though the use of name, value lists. The `normCompare.pl` Perl program was then overhauled to handle the new discrete characteristics.

Characteristic	Norm Sim
Destination IP packets	0.8182
Destination IP bytes	0.8182
Destination IP connections	0.75
Service packets	1.0
Service bytes	1.0
Service connections	1.0
Source IP packets	0.92
Source IP bytes	0.92
Source IP connections	0.9048
Source port packets	1.0
Source port bytes	1.0
Source port connections	1.0
TTL	1.0

Table D.12: The normalized similarities of discrete characteristics for the first run of `normCompare.pl` on basecase 1.

D.2.2 Dealing with IP address differences

On our first run of the new `normCompare.pl` program, we got the normalized similarities for basecase 1 shown in Table D.12. Considering that all the normalized similarities between the two traces in basecase 1 are supposed to be 1.0, these results give us pause. We quickly see that the problems are with all the “Source IP” and “Destination IP” characteristics, which makes sense given that the only differences between the two traces is the exact IP addresses on the local network.

Given this we tried adding an optional argument to `normCompare.pl`, which allows the user to specify the local network. If provided, it removes any hosts on that network from the “Source IP” and “Destination IP” characteristic measurements. Since the argument is optional, it allows a network manager who is interested in network comparison for looking at how their network evolves over time to include the local hosts. It is not helpful, however, if we want to know how two different networks differ, when the only difference is that one has one host that produces the same type and amount of traffic as two hosts on the other network.

Another issue that we should keep in mind as we look at this is that many large sites (such as `yahoo.com`) will operate a large number of servers on a single subnet, directing visitors to different ones by round-robinning the responses to DNS requests. Given that, there may be some advantage in lumping all requests to the same class-C network together into a single count. This does not, however, address the other problems, above.

We also considered taking the vector of measurements for those characteristics, ordering it from highest to lowest count, and comparing the discrete measurements that way. That, however, is essentially what we will be doing when we use the “Source IP” and “Destination IP”

characteristics in a continuous distribution, so we will wait for that then.

At first glance, what we want is a way to say, “On this network, treat the hosts as interchangeable, such that if I do not see a given host from the first trace in the second trace, but see one that acts just like it, treat them as the same host.” If we take one step back though, we must remember that the traces may be from two different networks, so what we really want to say is, “For each host on the first network, try to uniquely (1:1) map it to a host on the second network, and treat them as the same host.”

We implemented this idea in the `mapHosts.pl` Perl program, which takes two distribution files, extracts the discrete characteristics for each host, compares every host on the local network of the first trace to every host on the local network of the second trace (although this is an $O(n^2)$ operation, n is only 256), and writes out two new distribution files with unique names in place of each pair of IP address. The pairs are assigned from highest similarity to lowest, until all the hosts in one of the two networks are assigned. Similarity is defined as the mean of the similarities for those hosts for each discrete characteristic. This may mean that two very different hosts get assigned the same unique ID, which is fine because the radical differences between them will surface when they are compared in the `normCompare.pl` program. It may also mean that two hosts on the same network swap places between two traces. This is fine if we are only concerned with the aggregate activity on the network. If a network manager does want to take into account the hosts swapping activity profiles, then they should not run the `mapHosts.pl` script. When we ran basecase 1 through both `mapHosts.pl` and `normCompare.pl`, we got all 1.0 scores, as we should.

The other four basecases were also run through the `normCompare.pl` program, providing a much larger diversity of similarity scores than the count characteristics alone did. For basecases 2, 3, and 4, the use of `mapHosts.pl` did not make a difference, since all three of those basecases compare one network to itself. We did use `mapHosts.pl` on basecase 5, which provided a small score for some characteristics that otherwise would have been zero. We think this makes sense, because `mapHosts.pl` is intended to be used to compare different networks, which is effectively what basecase 5 represents. The normalized similarities are shown in Table D.13.

D.2.3 Non-linear versus linear weighting

Using the normalized similarities shown in Table D.13, and our original weights of (1.0, 1.0, 0.9, 0.8, and 0.1) we ran the `evalSimsNonlinear.pl` program, to find the non-linear factors shown in Table D.14. We immediately see that we have a lot of inverted parabolas. As before, we changed our goals to the mean of the normalized similarities for each basecase, resulting in new

Characteristic	Basecase 1	Basecase 2	Basecase 3	Basecase 4	Basecase 5
Bytes in	1.0	1.0	0.9994	0.9995	0.0083
Bytes out	1.0	1.0	0.9994	0.9506	0.0006
Connections in	1.0	1.0	1.0	0.9873	0.1642
Connections out	1.0	1.0	1.0	1.0	0.1596
Destination IP packets	1.0	1.0	0.6666	0.4224	0.0032
Destination IP bytes	1.0	1.0	0.5556	0.3987	0.0135
Destination IP connections	1.0	1.0	0.6666	0.4269	0.0100
Packets in	1.0	1.0	0.9989	0.9978	0.1327
Packets out	1.0	1.0	0.9992	0.9967	0.1101
Service packets	1.0	1.0	0.8916	0.7178	0.0
Service bytes	1.0	1.0	0.8627	0.7617	0.0
Service connections	1.0	1.0	0.8885	0.7206	0.0
Source IP packets	1.0	1.0	0.7143	0.4981	0.0001
Source IP bytes	1.0	1.0	0.6667	0.5185	0.0035
Source IP connections	1.0	1.0	0.7143	0.4994	0.0037
Source port packets	1.0	1.0	0.8942	0.7531	0.0
Source port bytes	1.0	1.0	0.875	0.7941	0.0
Source port connections	1.0	1.0	0.8944	0.7531	0.0
TTL	1.0	1.0	0.7962	0.6412	0.1397

Table D.13: The normalized similarities of count and discrete characteristics for basecases 1 through 5, using `mapHosts.pl`.

Characteristic	b	c
Bytes in	12.0834	-11.1614
Bytes out	40.3593	-39.8838
Connections in	0.5335	0.3961
Connections out	0.5700	0.3550
Destination IP bytes	2.8739	-1.9394
Destination IP connections	2.8249	-1.8574
Destination IP packets	2.7608	-1.8152
Packets in	0.7239	0.2021
Packets out	0.9012	0.0249
Service bytes	3.3716	-2.5284
Service connections	3.6850	-2.8744
Service packets	3.3848	-2.5436
Source IP bytes	2.7719	-1.8506
Source IP connections	2.6562	-1.7211
Source IP packets	2.7031	-1.7745
Source port bytes	3.7414	-2.9245
Source port connections	4.2742	-3.4987
Source port packets	3.7435	-2.9266
TTL	1.1334	-0.0830

Table D.14: The factors for the non-linear weighting of the count and discrete characteristics, using the goal similarities of (1.0, 1.0, 0.9, 0.8, and 0.1).

goals of (1.0, 1.0, 0.8465, 0.7283, 0.0394). Rerunning the `evalSimsNonlinear.pl` program gives us the new factors shown in Table D.15. While that corrected some of the curves, and generally reduced the magnitude of most of them, the curves are still generally outside our acceptable bounds.

At this point, we have come to the realization that the problems with non-linear weighting, first noted with the “Bytes in” and “Bytes out” characteristics, are not limited to mere outliers that should be discarded. Instead, the behavior of these characteristics are fundamental to the trace comparisons being performed. We considered dropping the weighting step from the methodology altogether; however we can not ignore the belief that some characteristics are going to be more descriptive of the similarities between traces than others.

Since the fundamental problem we seem to be have with the non-linear weighting is over-fitting of the data, we decided the more general linear weighting. For each characteristic, we will determine a weighting factor, wf :

$$wf = \frac{\sum_{i=1}^n \frac{norm_i}{goal_i}}{n}$$

where n is the number of baselines, $norm_i$ is the normalized similarity of the i^{th} baseline, and $goal_i$ is the goal value for the i^{th} baseline. We reset the goals to (1.0, 1.0, 0.9, 0.8, 0.1) for the first five baselines, and ran the `evalSimsLinear.pl` Perl script to calculate these weighting

Characteristic	b	c
Bytes in	4.7464	-3.8535
Bytes out	14.2862	-13.5364
Connections in	0.0938	0.8058
Connections out	0.1241	0.7696
Destination IP bytes	2.2859	-1.3170
Destination IP connections	2.3284	-1.3356
Destination IP packets	2.1938	-1.2160
Packets in	0.2018	0.6939
Packets out	0.2848	0.6106
Service bytes	1.7951	-0.8655
Service connections	1.7920	-0.8665
Service packets	1.7914	-0.8617
Source IP bytes	2.0998	-1.1353
Source IP connections	2.0276	-1.0508
Source IP packets	2.0472	-1.0771
Source port bytes	1.8182	-0.8962
Source port connections	1.8878	-0.9753
Source port packets	1.8186	-0.8967
TTL	0.7158	0.3439

Table D.15: The factors for the non-linear weighting of the count and discrete characteristics, using the goal similarities of (1.0, 1.0, 0.8465, 0.7283, 0.0394).

factors. The results are shown in Table D.16.

We can then calculate the scaled similarity value as a weighted arithmetic mean:

$$\frac{\sum_{i=1}^m value_i w f_i}{\sum_{i=1}^m w f_i}$$

where m is the number of characteristics. We computed this using the weights given above, and got the results in Table D.17.

Certainly, these are great results; however, looking at the weights, we see that the weights for some of the characteristics were above 1.0, however 1.0 should mean that the measure of that characteristic gives exactly the expected value in each trace pair. So if 1.0 is perfect, how can a value be above that? If we look back at our formula for calculating the weighting factors, we see that this is possible if the normalized similarity is higher than the goal. In other words, if a characteristic reports a measure of 1.0 for an expected value of 0.5, that will carry twice the weight of the characteristic that correctly reports 0.5. As such, we will modify the formula to calculate the expected weight.

$$w f = \frac{\sum_{i=1}^n 1 - |goal_i - norm_i|}{n}$$

After making this correction, we get the weights shown in Table D.18. We then re-ran the `evalCalcLinear.pl` script and got the calculated weights shown in Table D.19, which seem quite reasonable.

Characteristic	wf
Bytes in	0.8886
Bytes out	0.8610
Connections in	1.1975
Connections out	1.1914
Destination IP bytes	0.6601
Destination IP connections	0.6501
Destination IP packets	0.6748
Packets in	1.1369
Packets out	1.0914
Service bytes	0.7776
Service connections	0.7821
Service packets	0.7776
Source IP bytes	0.6834
Source IP connections	0.6847
Source IP packets	0.6911
Source port bytes	0.7870
Source port connections	0.7930
Source port packets	0.7870
TTL	1.0166

Table D.16: The factors for the linear weighting of the count and discrete characteristics, using the goal similarities of (1.0, 1.0, 0.9, 0.8, 0.1).

Basecase	Goal	Calculated
Basecase 1	1.0	1.0
Basecase 2	1.0	1.0
Basecase 3	0.9	0.8692
Basecase 4	0.8	0.7656
Basecase 5	0.1	0.0515

Table D.17: The expected and calculated similarities for the first five basecases using the count and discrete characteristics and the first linear weighting formula.

Characteristic	wf
Bytes in	0.9219
Bytes out	0.9301
Connections in	0.9297
Connections out	0.9281
Destination IP bytes	0.8584
Destination IP connections	0.8336
Destination IP packets	0.8607
Packets in	0.9341
Packets out	0.9388
Service bytes	0.9619
Service connections	0.9649
Service packets	0.9618
Source IP bytes	0.8825
Source IP connections	0.8777
Source IP packets	0.8835
Source port bytes	0.9695
Source port connections	0.9738
Source port packets	0.9695
TTL	0.9395

Table D.18: The new factors (using the fixed formula) for the linear weighting of the count and discrete characteristics, using the goal similarities of (1.0, 1.0, 0.9, 0.8, 0.1).

Basecase	Goal	Calculated
Basecase 1	1.0	1.0
Basecase 2	1.0	1.0
Basecase 3	0.9	0.8513
Basecase 4	0.8	0.7347
Basecase 5	0.1	0.0398

Table D.19: The expected and calculated similarities for the first five basecases using the count and discrete characteristics and the fixed linear weighting formula.

These weighting factors are actually quite interesting. Of our six count and 13 discrete characteristics, the highest weight is 0.9738, and the lowest is 0.8336. This means that some characteristics (Source port connections, in this case) have a very close correspondence to our intended similarity, and the characteristics that are most divergent from the intended similarity (Destination IP connections, in this case), are still fairly close. One might argue that this is an artifact of our artificially constructed trace pairs, as they were intended to have a given level of similarity. The fact that this similarity holds across all of our characteristics so far is indicative that we are on the right path.

Appendix E

More problems with linear scaling

The previous appendix covered the original problems we had with the count and discrete metrics. When we revisited the singular and discrete metrics (chapter 5), we ran into an additional problem with scaling, which we cover here in the interest of reducing clutter in that chapter.

As discussed in appendix D, we originally tried to use a quadratic scaling method, and eventually found that linear weighting factors worked better. These weighting factors were, calculated as

$$wf = \frac{\sum_{i=1}^n 1 - |goal_i - norm_i|}{n}$$

where n is the number of basecases, $norm_i$ is the normalized similarity of the i^{th} bas testcase, and $goal_i$ is the goal value for the i^{th} bas testcase. When we applied this to the normalized similarities shown in Table 5.1, we get the weighting factors shown in Table E.1.

One of the things we notice about these weights are that they are in the range [0.76, 0.98), which would imply that all of them carry some significant degree of importance in the calculation of the scaled similarity, however if we look again at the normalized similarities, we see that some of the metrics do not look useful for predicting the goal values. In particular, the SYN-ACK rate has a value of 1.0 for every entry, meaning it does not provide any information regarding how the five basecases are different.

Given this, it is not too surprising when we calculate the scaled similarity value as a weighted arithmetic mean:

$$\frac{\sum_{i=1}^m value_i wf_i}{\sum_{i=1}^m wf_i}$$

where m is the number of characteristics. Using the weights shown in Table E.1, we get the results shown in Table E.2.

Characteristic	wf
Bytes in	0.9232
Bytes out	0.9302
Connections in	0.9421
Connections out	0.9296
Packets in	0.9331
Packets out	0.9382
SYN-ONLY rate	0.7863
SYN-ACK rate	0.76
Idle connection rate	0.92
Half-open connection rate	0.7835
Service bytes	0.9612
Service connections	0.9731
Service packets	0.9614
Source port bytes	0.9687
Source port connections	0.9583
Source port packets	0.9691
TTL	0.8545

Table E.1: The weighting factors for the linear weighting of the singular and discrete characteristics, using the goal similarities of $\{1.0, 1.0, 0.9, 0.8, 0.1\}$.

Basecase	Goal	Calculated
Basecase 1	1.0	1.0
Basecase 2	1.0	1.0
Basecase 3	0.9	0.9427
Basecase 4	0.8	0.8816
Basecase 5	0.1	0.2046

Table E.2: The goal and calculated similarities for the first five basecases using the singular and discrete characteristics.

Linear Regression Model

Expected =

```

0.057 * Connections in count +
0.0434 * Bytes in count +
0.0887 * Bytes Service discrete +
0.0499 * Packets in count +
0.0555 * Bytes out count +
0.049 * Packets out count +
0.0798 * Packet Source port discrete +
0.2343 * Half-open connection rate ratio +
0.0428 * Idle connection rate ratio +
0.0801 * Bytes Source port discrete +
0.0504 * Connections out count +
0.2172 * Packet TTL discrete +
0.0665 * Connection Source port discrete +
0.0868 * Packet Service discrete +
0.0586 * Connection Service discrete +
-0.26

```

Figure E.1: Results of Weka's Linear Regression on our normalized similarities from Table 5.1, with the goal values of $\{1.0, 1.0, 0.9, 0.8, 0.1\}$.

Considering the problems with the SYN-ACK rate, we hypothesized that dropping it (or setting the scaling weight to 0) would provide better performance. A quick test gives us scaled similarities of $\{1.0, 1.0, 0.9397, 0.8755, 0.1636\}$, which is indeed slightly better, but not by much. It seems that we need to apply a weight factor to each metric that takes into account not only its mean accuracy in predicting the desired values, but also if the difference between the values is to the same degree as the difference between the expected values; essentially we want the information gain that the given metric provides. The SYN-ACK rate does not provide any information gain, hence why it should have a zero weight. After some consideration of how to accomplish this, the thought occurred to us that linear regression does exactly this. So we took our normalized similarities, fed them into the Weka (University of Waikato 2007) Linear Regression algorithm, and got the output shown in Figure E.1.

Basecase	Goal	Calculated
Basecase 1	1.0	1.0
Basecase 2	1.0	1.0
Basecase 3	0.9	0.9000
Basecase 4	0.8	0.8000
Basecase 5	0.1	0.1000

Table E.3: The goal and calculated similarities for the first five basecases using the singular and discrete characteristics with the weights found by linear regression.

These weights look a lot better, as one will note that they do not include either SYN-ACK rate or SYN-ONLY rate in the equation, and the other weights range from 0.0428 to 0.2343, indicating that the different metrics do indeed have a factor of almost six difference in the result. The only part of this approach that gives us pause is the inclusion of a y-intercept – this is essentially a constant correction factor, which we can not provide a logical justification for. It also gives us pause to consider if we could provide any normalized similarity values that would result in a scaled similarity value outside the range [0..1]. If we consider the case where all the metrics have the maximum normalized similarity of 1.0 and do the calculation, we see that we do indeed end up with 1.0. This should not be surprising since we had just such a case in our basecases (two, actually), and the LR model fit the basecases with no error. On the other hand, if all the metrics are 0.0, then the scaled similarity will be -0.26 . While this is somewhat troublesome, we must first consider that all the metrics can not, in fact, be zero, unless one trace file had no traffic in it. Furthermore, any case that causes our scaled similarity to fall outside the range [0..1] must be considered a case that we did not account for when calibrating the metrics for a particular instantiation of the methodology, and hence a flag that we must recalibrate and rerun our comparisons. We can find an analogy with any piece of measuring equipment; for example, a thermometer that is calibrated between $273K$ and $373K$ may give an erroneous measurement of $-0.02K$ when used at extremely low temperatures.

We will move forward using the Linear Regression Model to find both our weights, and our y-intercept. We created the `parseModel.pl` script to convert the textual output of Weka's Linear Regression Model to a form we can easily use, and the `calcSimLinear.pl` script to use those weights on the normalized similarities to generate the calculated scaled similarities, shown in Table E.3.

Appendix F

Modeling Interarrival time of packets

This appendix details our work on developing our sorted continuous modeling method based on our sample Packet Interarrival time data. Details on most of the other continuous characteristics are in Appendix H. The end result of this work is the ordered continuous modeling method presented in chapter 7. The definitions for any terms used here can be found in section 7.2.1.

Much research has been done on the modeling of the time delta between packets (see chapter 2), as this serves as a useful measure of the latency in a network. This may be good in that a lot of our work is done for us, or it might be bad in that there is so much to consider. We will take the simplistic approach and summarize fifteen years of research as follows: *Packet interarrival times follow a power-law distribution*. This is easy to believe when one looks at figure Figure F.1.

Our problem occurs when we actually go to fit a power function to the data, as one can see in figure Figure F.2. Results are similar if we try to fit an exponential curve to the data. It almost looks as if the rise is too sudden and steep to capture with our available fitting algorithms.

As such, we decided to try to do the analysis by hand. As a first step, we took the \log_{10} of the delta, giving us figure Figure F.3. This is actually quite fascinating, as it shows a pattern in the data that one does not see when it is plotted out at a normal scale. We suspect that the stair-step appearance is the manifestation of a physical phenomenon, such as number of routers the packet must pass through, or queuing latency. It should be interesting when we get to the higher-order characteristics to see if there is a correlation between these plateaus and

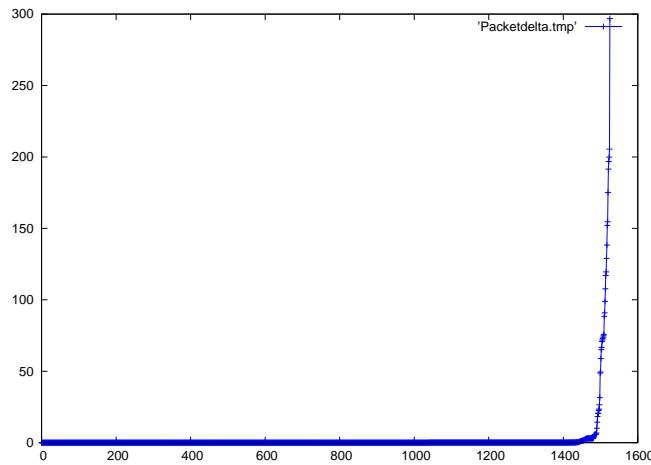


Figure F.1: Plot of Interarrival time of packets values for one hour of sample data.

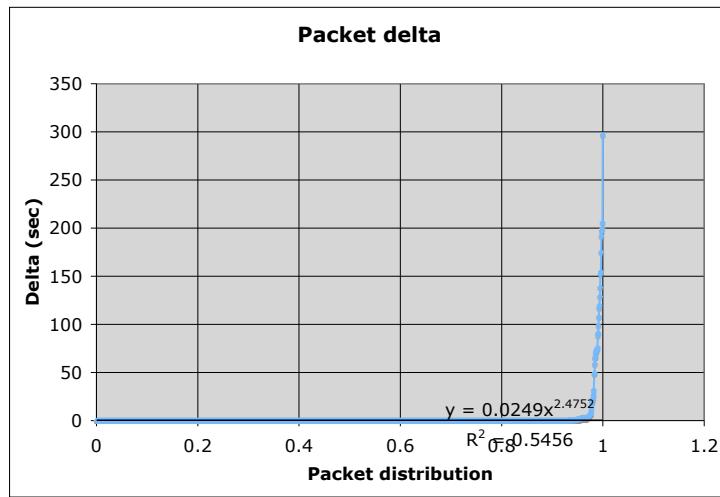


Figure F.2: Plot of Interarrival time of packets values for one hour of sample data with a power function fitted to the data.

other characteristics.

For the moment, we are left to try to figure out how to model this log plot. If we fit a sixth-order polynomial, we get the plot shown in figure Figure F.4. While the actual fit (R^2) is 97%, a visually analysis of the curve indicates that it is not actually that good of a fit. In particular, it fails to capture the three plateau we see. Furthermore, if we use the given equation (with the given four digits of precision) and try to recreate the log curve, it actually moves in the opposite (negative) direction as x increases. This indicates a numeric instability, or some type of chaotic property to the equation. To address these problems, we could build a piece-wise function, made up of three third order polynomials, as seen in figures Figure F.5 through Figure F.7. While there is some overlap between the three curves in order to capture the behaviour of the curves, one can easily imagine the stitched together function which would switch between pieces halfway through the overlap, as seen in figure Figure F.8. If we then use these \log_{10} to generate packet deltas and plot them out, we get the plot shown in Figure F.9.

This seems to be a reasonable approximation of the data. We are left with two concerns. First of all, we would like the implementation of our methodology to be, to a large degree, automatable, and having an algorithm that generates piece-wise functions of n-degree polynomials seems difficult at first glance. Second, the true test of the usefulness of a given continuous function representation will only come when we try to find the similarity between two curves.

To that end, we started to think about how two of these curves could be compared. The most straightforward approach would be the absolute difference in area under the curve. That is, in discrete terms, for every point on the curve, we take the absolute value of the difference between that value and the corresponding point on the other curve. The sum of these differences is divided by the number of difference calculations to give us the absolute difference in area under the curve. Consider figure Figure F.10 – the absolute difference in area under the curve is the total area between the two curves. Note that both the x and y values have been scaled to fit in the range [0:1] – the x values separately, based on how many measurements were made, and the y values by the maximum value between the two datasets. In doing so, the similarity measure is constrained to [0:1].

If both curves are represented as continuous functions, then sampling the curves to find the absolute difference in area between them only gives an approximate solution – a true solution would need to subtract the functions from one another and use that to calculate an exact solution. But since our data is already discretized, we can just calculate a sufficiently close solution from that, in which case, building a continuous function does not really buy us anything¹. To do this, we take each discrete value from one of our datasets and compare it to the corresponding

¹Other than paper fodder, of course.

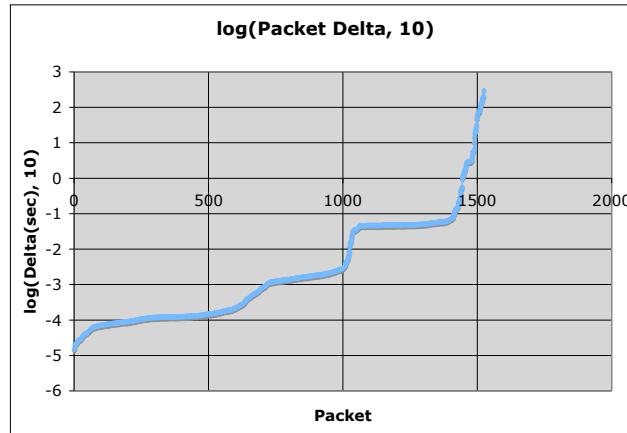


Figure F.3: Plot of \log_{10} (Packet Interarrival Time) values for one hour of sample data.

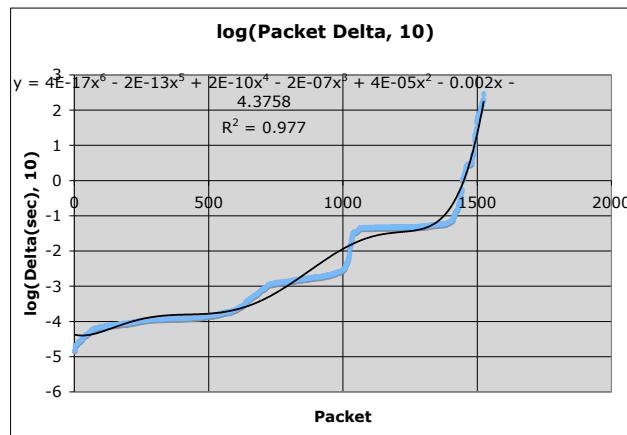


Figure F.4: Plot of \log_{10} (Packet Interarrival Time) values for one hour of sample data with fitted sixth-degree polynomial.

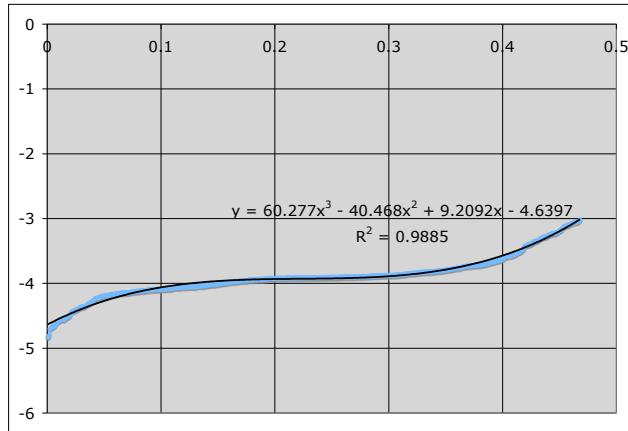


Figure F.5: First piece of fit curve: third-degree polynomial covering about 45% of the data.

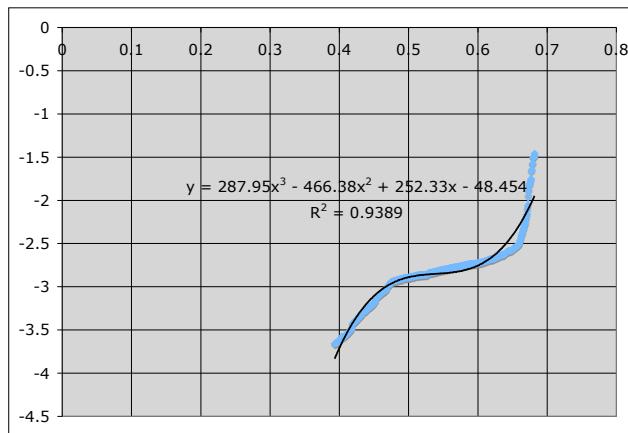


Figure F.6: Second piece of fit curve: third-degree polynomial covering about 30% of the data.
Note overlap with first piece.

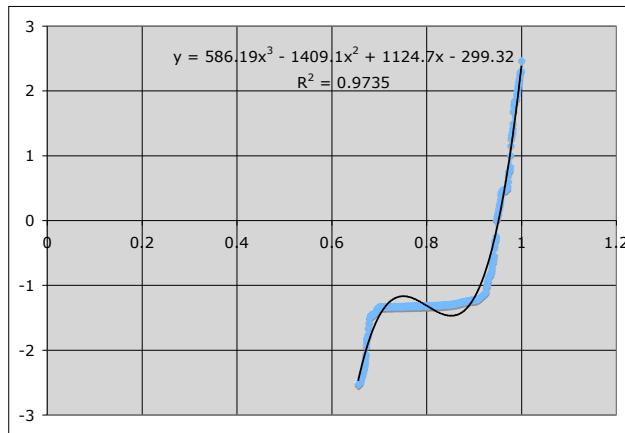


Figure F.7: Third piece of fit curve: third-degree polynomial covering about 35% of the data.
Note overlap with second piece.

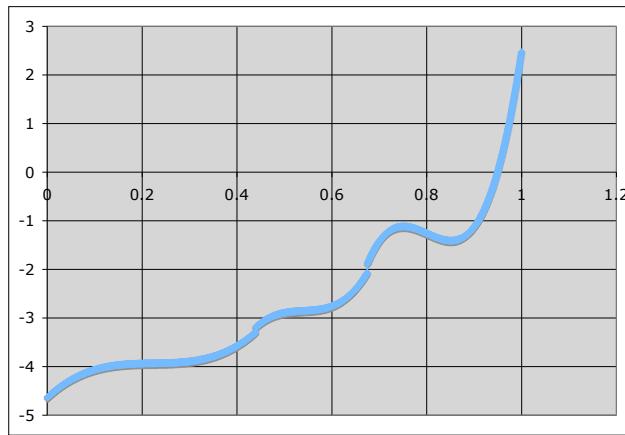


Figure F.8: Piece-wise function fitted to \log_{10} of packet deltas.

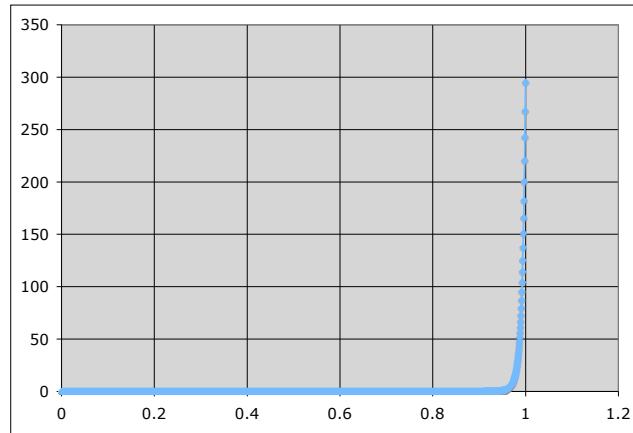


Figure F.9: Generated packet interarrival times using \log_{10} values seen in figure Figure F.8.

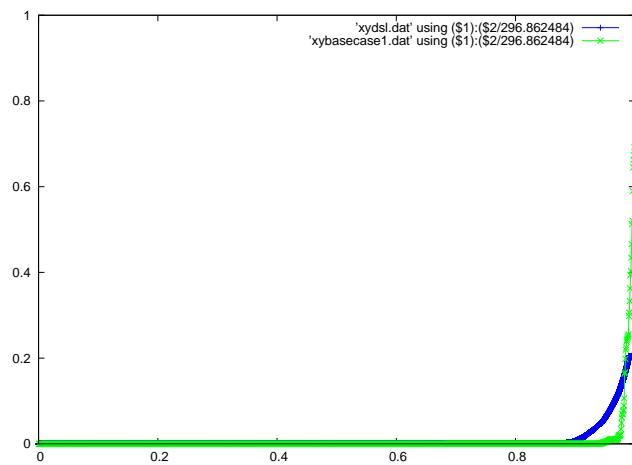


Figure F.10: Plot of interarrival times of packets for both one hour of sample data (green line) and DSL data (blue line). The x-axis of each dataset was scaled separately so that both would fit in the range [0:1]. The y-axis of both datasets were scaled together by the maximum value in both (296.862484) so that both would fit in the range [0:1] while maintaining their relative differences.

value (same x in the range [0:1]) in the other dataset. If there is not a corresponding value with the exact same projected x , we take the average of the values corresponding with the closest projected x 's on either side. This works sufficiently well because

- the distance between the x values is constant,
- we have a large number of values, even for a small dataset,
- the y values are monotonically increasing, and
- we loop over the smaller of the two datasets so the distance between the x values will always be larger than the x values of the dataset we are comparing to, meaning the same pair of projected x 's will not be used for successive averages.

When we do this, we find that the calculated similarity (one minus the difference) between the two curves in figure Figure F.10 is 99.2195%. Given how different the two curves look, this seems rather large. We must consider though that both curves only diverge in the last 10% of the data, and they have a similar kind of shape when they do diverge – so the total area difference between them really is something like 0.8%. This is a pretty small difference to use as the basis of a comparison between two networks though; especially when we consider that the blue line starts diverging from zero around 0.9 and the green line does not diverge significantly until 0.97 or 0.98.

In order to increase the differentiation value based on the shape of the curve, we try scaling the y values separately, so that both curves run up to one. The result can be seen in figure Figure F.11. When we calculate the similarity after scaling the curves this way, we get a similarity of 96.8311%, which better reflects the differences in shapes of the curves. Unfortunately, the approach discards any meaningful comparison of the actual latency values – if the latency of all the packets on one network is consistently double that of another network, this method will calculate that they are 100% similar.

We consider that part of our problem is that the comparison is being dominated by the large values to the right side of the graph. To address this, we turn again to the logs of the interpacket arrival times. If we plot out the \log_{10} of the two curves, using the same x scaling as before, we get the plot shown in Figure F.12. This produces a very nice similarity value of 33.4611%², until we realized that the value was not constrained to the range [0:1]. Sure enough, given the arrays $\{0.01, 0.02, 0.03\}$ and $\{1.0, 2.0, 3.0\}$, the method calculates a similarity value of -1 .

²The comparison based on the log values discards any zeros in the dataset.

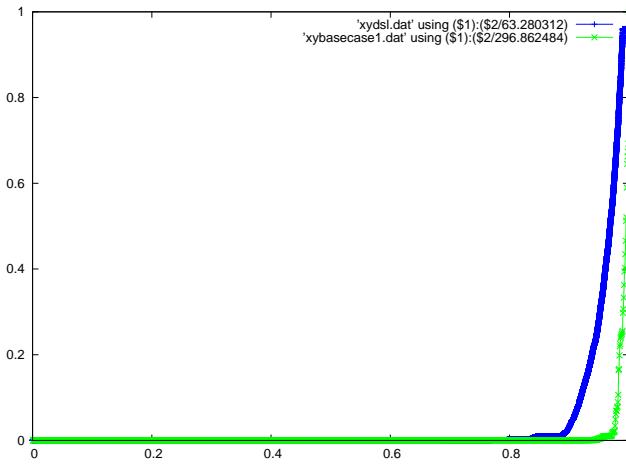


Figure F.11: Plot of interarrival times of packets for both one hour of sample data (green line) and DSL data (blue line). The x-axis of each dataset was scaled separately so that both would fit in the range [0:1]. The y-axis of each datasets was scaled separately (63.280312 for the blue line and 296.862484 for the green line) so that both would scale the full range of [0:1].

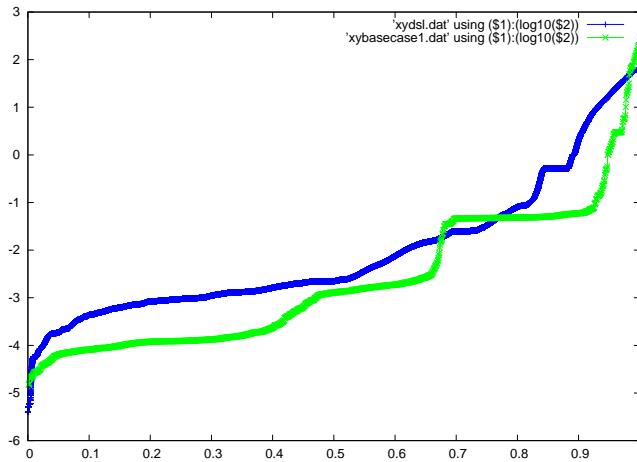


Figure F.12: Plot of the \log_{10} of the interarrival times of packets for both one hour of sample data (green line) and DSL data (blue line). The x-axis of each dataset was scaled separately so that both would fit in the range [0:1].

Once again, we scale the data, this time by subtracting the minimum log value from all the other log values so that the y values will start at zero, and dividing by the range of values, so that the y values will go up to one. Both datasets are scaled by the same amounts so that the relative differences remain. The result is shown in figure Figure F.13. Now the similarity value is a somewhat more reasonable 91.5458%.

What are we actually comparing here though? The scale of the two datasets when plotted side-by-side? If we use the example above, with the arrays $\{0.01, 0.02, 0.03\}$ and $\{1.0, 2.0, 3.0\}$, the method now calculates a similarity value of 19.2611%. While that is at least a valid value (unlike before), it is hard to justify an almost 20% similarity for two datasets that differ by two orders of magnitude. Given that we have now figured out how to obtain a y value for each dataset at numerous x values, why do not we just use the average of the same similarity calculation we are using for discrete data:

$$1 - \frac{|y_1 - y_2|}{y_1 + y_2}$$

If we do this for the arrays that differ by two orders of magnitude, we get the much easier to understand similarity value of 1.9802%, and for our two interpacket arrival curves, we get a similarity value of 41.7172%. We can even see how we came to this value by plotting out the similarity values across the $[0:1]$ range for x , as in figure Figure F.14. Note that the \log_{10} values of the two datasets are shown only so that the similarities and differences between the two datasets are more legible.

With this, we finally feel that we have a way of modeling the data such that we can calculate similarity values between sets of interpacket arrival times.

In considering the other continuous metrics that look promising to model the same way, it occurs to us that we became fixated on plotting out the values in sorted order, based on the prior work that had been done on packet latency and long tailed distributions. The downside of this approach is that it makes the independent and identically distributed (iid) assumption, which may not be the best approach; we may be better off with an approach that takes the order the values come in into account, such as time series analysis. We leave this for future research.

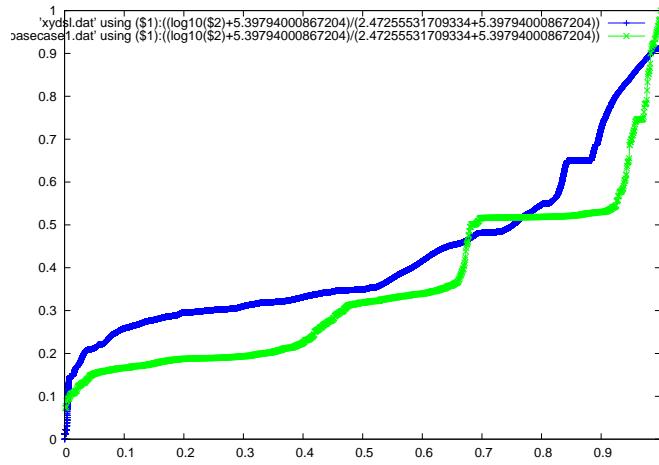


Figure F.13: Plot of the \log_{10} of the interarrival times of packets for both one hour of sample data (green line) and DSL data (blue line). The x-axis of each dataset was scaled separately so that both would fit in the range [0:1]. The y-axes of both datasets were scaled by the same amount so that they would both fit in the range [0:1].

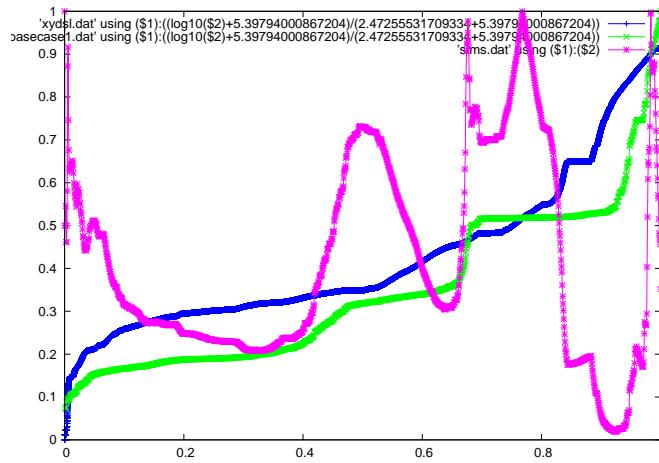


Figure F.14: Plot of the similarity values (purple line) between the interarrival times of packets for one hour of sample data (green line) and DSL data (blue line). The x-axis of the similarity values aligns with the scaling of the two datasets to both fit in the range [0:1]. The y-axis of the similarity values is true. The y-axes of the two datasets are shown as scaled \log_{10} so that the similarities and differences between them are more visible.

Appendix G

Modeling TTL

This appendix details our work on developing our ordered continuous modeling method based on our sample TTL data. Details on most of the other continuous characteristics are in Appendix H. The end result of this work is the ordered continuous modeling method presented in chapter 7. The definitions for any terms used here can be found in section 7.2.1.

Figure Figure G.1 shows the distribution of the time to live (TTL) values on one hour of sample data.

Given that the network we are looking at here consists of a single Linux host, and many of the hosts that it communicates with are Linux or BSD based, it comes as no surprise that the bulk of the observed TTL values are at and approaching 64. We also have the expected cluster of values approaching 128 from modern Windows-based systems. Finally we have a blip at 254, most likely from a router. There is also a tiny blip at 19 which we can not easily attribute.

Fitting a function to this will likely be tricky, primarily due to the lack of data-points. While the peak values are at and approaching 64, in actuality, we only saw six TTL values in this range, and three of them dominate the traffic seen; in other words, the curve is not very smooth. Further, we know that the blip at 254 is significant, but it will be difficult to capture such domain knowledge in a function fitting algorithm.

We tried fitting numerous polynomials to the data, up to a sixth order polynomial, which gave us the following

$$6 * 10^{-11}x^6 - 4 * 10^{-08}x^5 + 1 * 10^{-05}x^4 - 0.0015x^3 + 0.0767x^2 - 0.7282x$$

Even with such a high-order polynomial, we only achieved a fit (R^2) of 0.0414.

We thought that, perhaps a Fourier analysis of the data points could find the component frequencies that combine to form the “waves” we see in the data. While we were able to use a FFT

to find the component frequencies, the removal of the high frequencies quickly caused a divergence when we attempted to invert the FFT to reassemble the original data points, particularly with the large number of zero data points we have. This indicates that the data is not well represented as a waveform.

We considered the number of functions we have in our arsenal, and all seem fraught with two main problems:

1. Even looking at the data as a human, it does not look contiguous. At best, it looks like a couple different ranges of values with lots of zeros in-between.
2. Even within the (small) ranges where we have non-zero values, the values are not very smooth.

We considered that the first problem might be addressed by looking at TTL values as a piece-wise function, with different functions (or at least different factors of the same function) representing different segments of the total range between zero and 255. Even if we did that, we would be left with the second problem: patchy data. To address this, we deviated from the one-hour of CoLo data we have been looking at so far, and grabbed approximately 2.5 days of IPv4 traffic from the author's residential DSL connection. The plot of the TTL we obtained from this data is shown in figure Figure G.2. This data is actually much more instructive, as we notice a number of points about it:

1. There are a number of packets with TTL values below 40 – in particular four each with values of 1, 2, and 3. These are likely `traceroute` packets – the others are more difficult to attribute.
2. There are a large number of packets with values from 40 to 57. These are likely remote packets from Linux/BSD-type servers. There is no clear pattern in the distribution of these counts, most likely due to a combination of a different number of packets sent between each remote box and this network, and a potentially chaotic network distance between the two. The second-order characteristic of TTL per IP should be particularly interesting.
3. Most of the counts between 58 and 62 are zero, except for two packets with a value of 61. This indicates that there are five layers of network infrastructure between the given DSL network and most any servers that communicate with it.
4. The spike of values at 63 and 64 are most likely the author's BSD based machines and Linksys router.

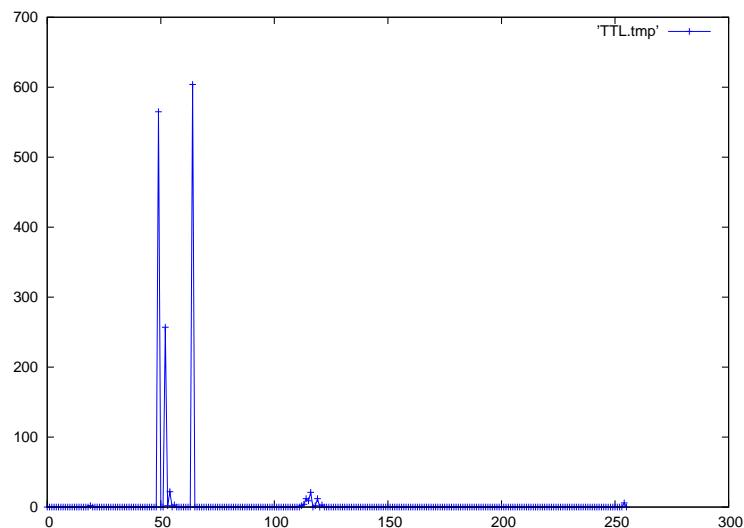


Figure G.1: Plot of time to live (TTL) values for one hour of sample data.

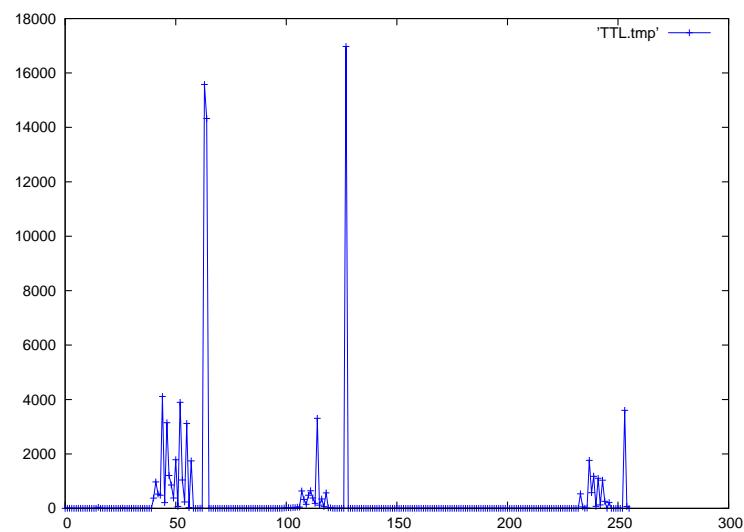


Figure G.2: Plot of time to live (TTL) values for approx 2.5 days of DSL data.

5. The values are once again flat until 97, where we see another possibly chaotic progression of values through 120, which are probably Windows based boxes. The wider range of TTL values here is likely due to the larger number of compromised Windows boxes around the globe used for scanning and worm propagation.
6. Here we apparently have six layers of infrastructure (TTL values 121 through 126) with a value of zero, except for another blip of 2 packets with a value of 125.
7. The large spike at 127 is undoubtedly the Windows machine on the author's network.
8. Values are zero again from 128 all the way to 233. This is interesting insofar as with all the noise on contemporary networks, it does not appear that any scanning or attack tools are using random TTLs.
9. We have a third collection of potentially chaotic values from 233 to 246, including a zero value at 245.
10. There are six zeros, from 247 through 252.
11. There is a spike at 253 and 254, which is most likely from the DSL provider's servers (particularly DNS) and routers. Here is where a second-order distribution of TTL per service would be interesting.

Considering the above, it does not seem like modeling the TTL values as a continuous function makes much sense. At the same time, it is not a purely discrete function: we would like to say that a network with 100 packets with a TTL value of 55 is more similar to a network with no TTL value 55 packets, but 100 TTL value 56 value packets than one with no TTL value 55 or 56 packets.

This insight led to the inspiration to use a windowed average. Some definitions are necessary to explain this:

- x is a given TTL value, and is our independent variable for plotting purposes
- $\text{count}(x)$ is the number of packets with the TTL value of x
- y is the dependent variable for plotting purposes
- f_x is some function that maps x to y

So, figure Figure G.2 represents $y = \text{count}(x)$, so $fx = \text{count}$. We want to create a windowed average such that

$$y = \frac{.25 * \text{count}(x - 2) + .5 * \text{count}(x - 1) + \text{count}(x) + .5 * \text{count}(x + 1) + .25 * \text{count}(x + 2)}{2.5}$$

The choice of a window size of five was based on the fact that each of the regions of activity were separated by at least five zero (or almost zero) values, so going two TTL values in either directions would keep the regions of activity from overlapping. The choice of .5 and .25 as weighting values was more arbitrary, based primarily on the intuition that the weights should follow a bell curve centered on x with $x \pm 3 == 0$.

We tried this for TTL values 40 to 57, which if we tried to fit a sixth-order polynomial to, only had a fit (R^2) of 0.18 . If we try to fit a sixth-order polynomial to the windowed average between 38 and 59 (we increase the x range to account for the window two values in either direction), we get a fit of 0.84 . It is worth noting that decreasing this to a fifth-order polynomial drops the fit to 0.755 . While this seems like a great advance, looking at the polynomial equations themselves is instructive. The polynomial equation for the original data is

$$0.0099x^6 - 0.6738x^5 + 17.317x^4 - 208.31x^3 + 1155x^2 - 2315.1x + 1506.7$$

(see upper plot, figure Figure G.3), and for the moving average data it is

$$0.0036x^6 - 0.2822x^5 + 8.4562x^4 - 120.77x^3 + 818.27x^2 - 2102.6x + 1568.4$$

(see lower plot, figure Figure G.3).

What is so interesting here is that, while the exact factors differ, each of the polynomial factors is within an order of magnitude. This makes sense when one thinks about it: the polynomial fitting function is essentially looking for the average between the values to get the best fit. One might be inclined to just discard the moving average approach based on this; however, we are left to wonder if other approaches, such as Fourier analysis will be as forgiving, or if feeding in a smoother curve will produce better results. We should note that (Barford et al. 2002) used a wavelet-based methodology (similar in principle to an Fourier transform) to time-based data and noted, “due to its high approximation order, our system cannot capture accurately sharp discontinuities in the data”, particularly as we see such discontinuities, such as between TTL values 64 and 65.

The other change that we made when building the above polynomials is that we only looked at a slice of the data where the values were non-zero. Again, we are left to consider that it may be best to consider TTL value counts as a piece-wise function.

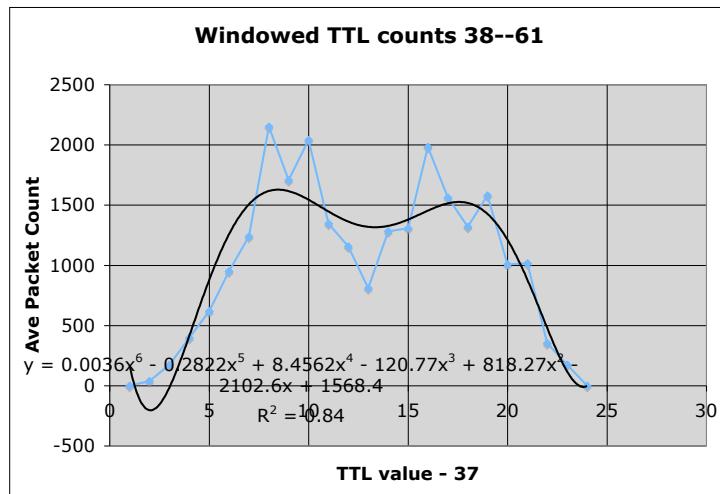
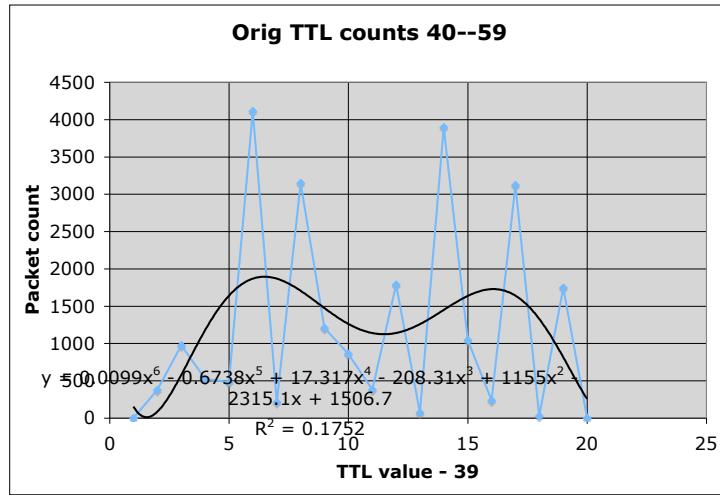


Figure G.3: Plot of original time to live (TTL) values 40–59 (above) and plot of windowed average time to live (TTL) values 38–61 (below), both with a fitted sixth-order polynomial

We decided to try the two approaches (raw values versus a windowed average) using the same slice of data with a FFT, and the results were particularly instructive.

The upper-left plot in figure Figure G.4 is an inverted FFT of the spectrum of 11 complex values¹ we get when we run a FFT on the packet counts for TTL values 40 through 59. These values are equal to the original packet counts. The successive plots, across and down, in the figure show what happens to the inverse FFT as we successively remove the highest frequency component. The last plot, in the lower-right corner, is the inverted FFT made up of only four spectral (complex) values. Note the similarity in this plot to the fitted sixth-order polynomials in figure Figure G.3.

Given this, figure Figure G.5 – which represents the same progression in successively removing high frequency components from the spectrum² – should come as no surprise. Of course, the progression is much less dramatic as the plot of the windowed average already visibly has the pattern of the ultimately inverted FFT of the compressed spectrum.

We are left with two conclusions at this point to drive our work forward:

1. The moving average only smoothes the data in a way analogous to the continuous functions.
2. Both the polynomial fit and Fourier analysis methods allow us to build a continuous function that is an approximation for the non-zero regions of data. Further analysis could yield others, however it is doubtful they will provide any benefit over these two.

So, what is the path forward? We must consider what the purpose of modeling the data as a continuous function is: to provide a better similarity metric than we can when we only look at the data discretely. Particularly for TTLs, the use of a continuous function does not provide any predictive value. If we were to use either a polynomial or an inverted FFT of the low-order spectral elements, how would these be used for comparison? We might use the difference between the curves, but one can imagine two sets of data producing the same approximation (at least for one segment of the curve) for different values.

Perhaps we could just consider the raw data points themselves as a curve and look at the area of difference between the two curves. This will address the above mentioned problem with providing a higher similarity value if one trace has packets with an adjacent TTL value versus another trace that does not. An example is illustrated in figure Figure G.6, if we have one trace (ex1) with packet counts 0, 10, 0, 0 (solid red line), a second trace (ex2) with packet

¹We should be able to represent the 20 real values with a spectrum of 10 complex values, however FFTs can act odd when run on an array of values that is not of a size that is a power of two, and it was not important enough to us to massage the values.

²Note that the plots with 10 and 7 spectral elements have been omitted for space and the fact that they looked nearly identical to the previous plots.

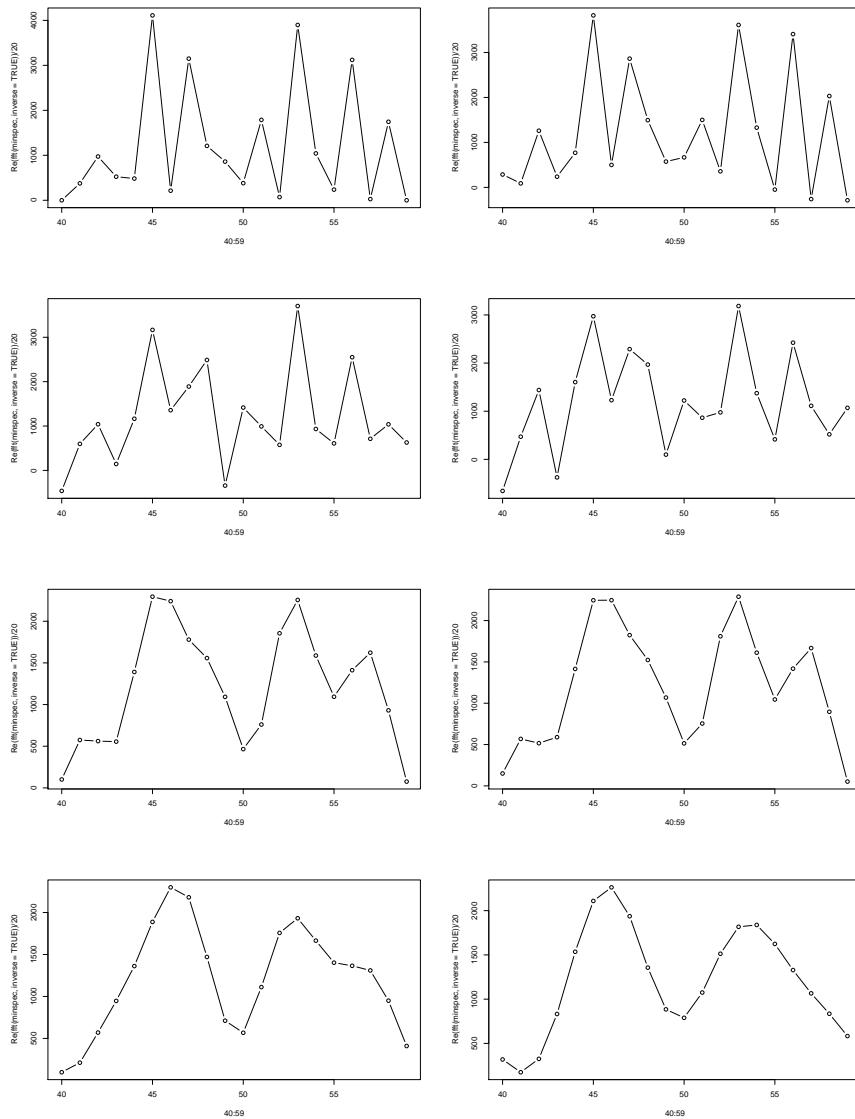


Figure G.4: Inverted FFT with 11 (upper-left) to 4 (lower-right) spectral (complex) values of original TTL values 40–59 – same as original values

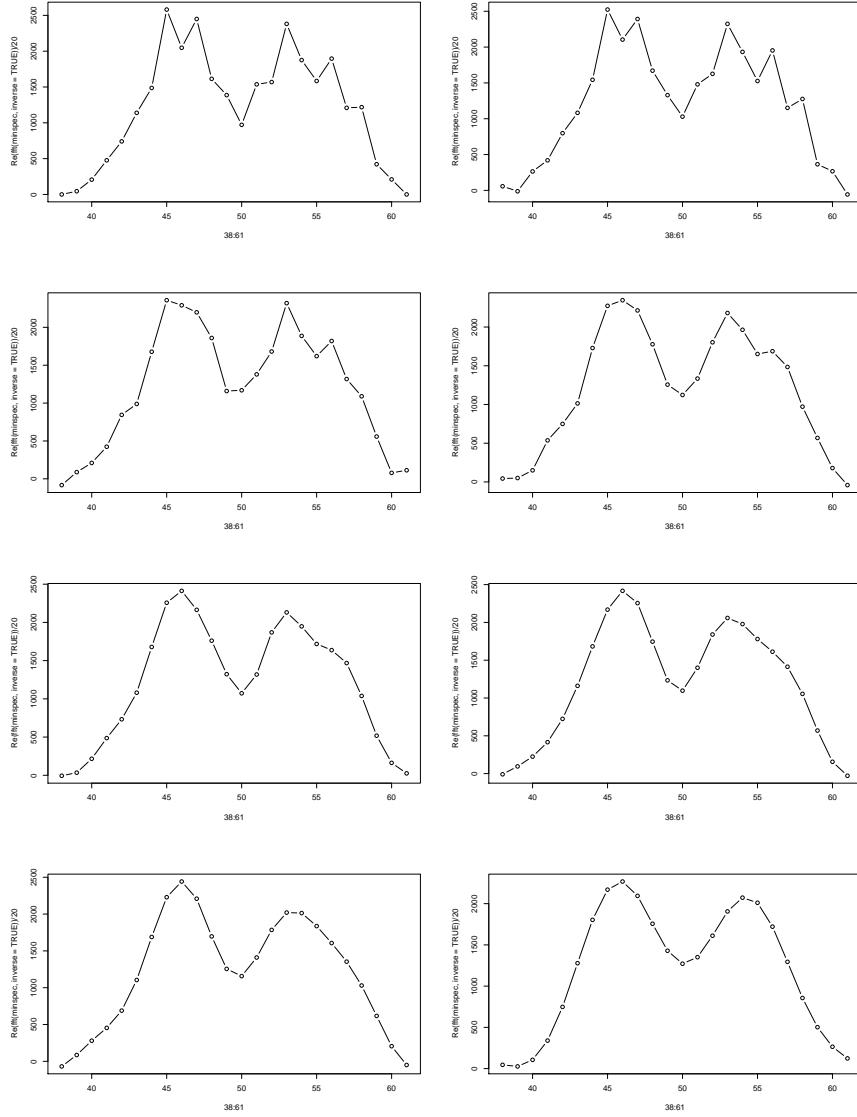


Figure G.5: Inverted FFT with (from upper-left to lower-right) 13, 12, 11, 9, 8, 6, 5, and 4 spectral (complex) values of windowed average TTL values 40–59; the plots with 10 and 7 spectral values have been omitted for space, and because they look almost identical to the plots with 11 and 8 spectral values, respectively

counts 0, 0, 10, 0 (dashed green line), and a third trace (ex3) with packet counts 0, 0, 0, 10 (dotted purple line), the common area between the first two traces will be 2.5, which is the area under both the solid red and dashed green lines. There is no similarity between the first and third traces though, as one can see that there is no area that is under both the solid red and the dotted purple lines.

This approach allows us to compare TTL distributions taking into account that adjacent TTL values are related, but not equal – hence they affect the similarity, but not at an equal level. Domain knowledge tells us that TTL values that are two apart are also related, but less so, and fall off to being mostly unrelated with a difference of three. The above approach only captures similarity between immediately adjacent TTL values. If we return to our moving average approach (figure Figure G.7), we may be able to obtain what we desire. Here we can see that the intersection in area between the first two traces is 13.75, versus 8.75 for the shared area between the first and third traces. We can also just apply our similarity formula for discrete values, which produces a similarity of 0.66 for the first two traces, and 0.45 for the second two traces. This sounds like just the scale of similarity values we are looking for, and allows us to leverage the discrete similarity code we already have.

In summary, in our first attempt to model data as a continuous function, we ended up just substituting a moving weighted average, and using our existing discrete similarity metric.

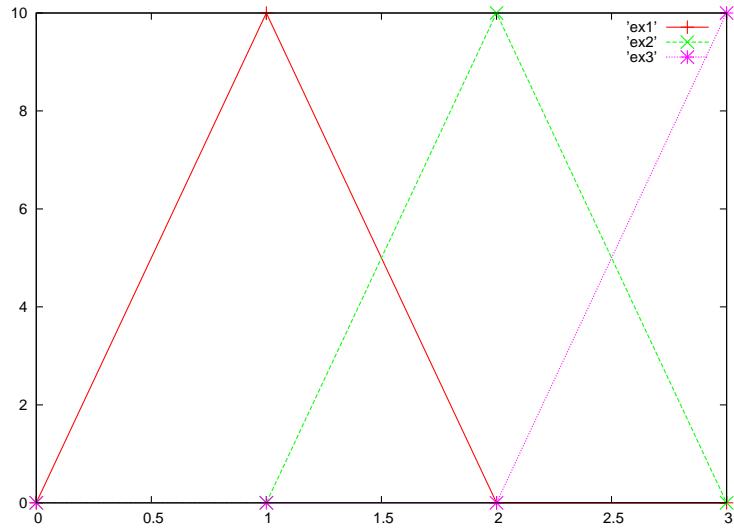


Figure G.6: Illustration of how raw data can be used to find difference in area under the curve

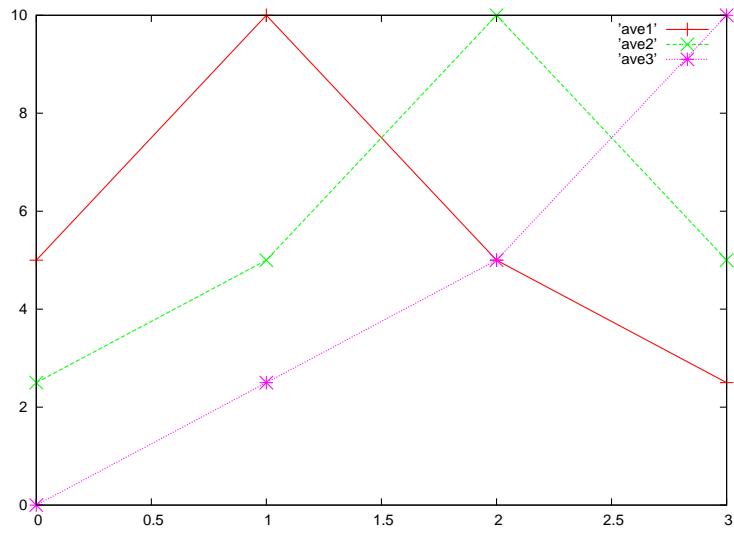


Figure G.7: Illustration of how raw data can be used to find difference in area under the curve

Appendix H

Details of Initial Continuous Characteristic Tests

This appendix contains additional tables and figures for chapter 7 which were not directly necessary for the accompanying discussion in that chapter. In particular, it provides an in-depth analysis and discussion of all the continuous characteristics themselves.

H.1 Modeling continuous distributions

This section will look at how we have chosen to model the 83 continuous characteristics. It relies heavily on the definitions provided in section 7.2.1. Graphs were constructed using the `genContin.pl` script, written by the author. By necessity, we begin to consider how such modeling impacts our similarity functions, however we save the formal calculations of similarity for section 7.3. We divide the section into subsections based on the divisions of continuous characteristics used in appendix B.

H.1.1 Per packet

Our per packet metrics are those metrics that we can extract from a stand-alone packet without any context (surrounding packets or connection information). We started with two of these – packet interarrival time and TTL – as representative examples of sorted and ordered continuous characteristics, respectively. As such, they end up setting the stage for the rest of our continuous characteristics.

Interarrival time of packets

Appendix F contains all the details of our work on the interarrival time of packets.

TTL

Appendix G contains all the details of our work on TTL modeled as a continuous characteristic.

Packets per seconds after the minute

As shown in Figure H.1, we have plotted out the number of packets that arrived for every second after the minute. That is, if one packet arrived at 08:01:03 and one packet arrived at 08:05:03, we say that two packets arrived at three seconds after the minute. Our intuition is that this will not be a worthwhile metric; we will include it anyway and determine what value it provides when calculating the weights. Our intuition that this metric is not very useful is based on the expectation that we expect to see a rather consistent level of traffic – that is, over a good sized dump of network traffic, we do not expect to see a significant difference in traffic based on the second after the minute; for example, an hour of network traffic will contain 60 samples for every second, each a minute apart (by definition). We expect this to result in a fairly flat line; Figure H.1 tells a different story, however. This is likely because our intuition is based on large networks, and this is a lightly used network with a single host, so we actually do see significant differences between the seconds. Some seconds have no activity; there was at least one burst of activity, and possibly two or three, between 18 and 25 seconds; and the activity we see between 32 and 46 seconds was likely a single large flow, such as a file transfer. As such, this may actually be a very useful for differentiating heavily and lightly used networks by something other than just the packet or connection counts.

How do we actually go about doing such a comparison? The intuitive way that we plotted this out, using the actual value of the second after the minute as the independent variable ordered along the x-axis, seems to make sense: we can see how it changes from second to second, and we see that there are patterns that emerge over multiple seconds. At the same time, the plot is not a smooth curve: we actually have multiple “islands” of activity. This is very similar to the issues we faced with the TTLs, so we will use the same approach for comparison: a weighted average to account for activity in adjacent seconds, and the ability to leverage our discrete comparison code.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of packets between the two traces.

Packets per minutes after the hour

Figure H.2 shows the packets per minute after the hour. This is a metric that we intuitively expect to be useful due to the combination of automated processes (cron jobs, checking pop mail every 15 minutes, auto-refresh on web pages, etc) and sociological reasons (checking a website at the top of every hour, for instance). Our plot here is interesting, as it shows a clear pattern of a burst of packets, followed by a slightly smaller burst of packets three minutes later; this pattern is repeated every ten minutes, for a total of six times over the course of the hour. Looking at the data, we see this is actually two separate processes – the first is Tor traffic, either doing something specific to Tor (such as rekeying activity), or tunneling traffic from an automated process. The second is SSH activity, most likely tunneling traffic from an automated process. Both processes happen to run on ten minute cycles.

Granted, this plot is actually rather clean because it is only looking at the minute after the hour for one hour of data – we are not seeing the combination of activity from multiple hours. Despite this, we feel confident that the proper way to compare this data is the same way we did TTLs and packets per second after the minute: do a discrete comparison based on a weighted average of the values.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of packets between the two traces.

Packets per hours after midnight UTC

Since we are only looking at one hour of sample data, Figure H.3 looks really boring. This is a measure that should be much more useful when looking at a whole day or more of data, where we expect to see typical daily patterns. For workplaces, we expect to see the bulk of activity between 08:00 and 17:00, with peaks at the start of the day and again right after lunch (Floyd and Paxson 2001; Barford, Kline, Plonka, and Ron 2002). For residential networks, we expect to see most of the activity in the evening, between the time people get home from work and when they go to sleep. The desire to plot data out by hours after midnight UTC is to be able to compare networks in different timezones – this is more useful for comparing external (incoming) traffic, which does not likely care what timezone the servers on this network are in. This is true for both benign traffic, such as external people accessing a webserver on this network, and for malicious traffic, such as the worm du jour.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of packets between the two traces.

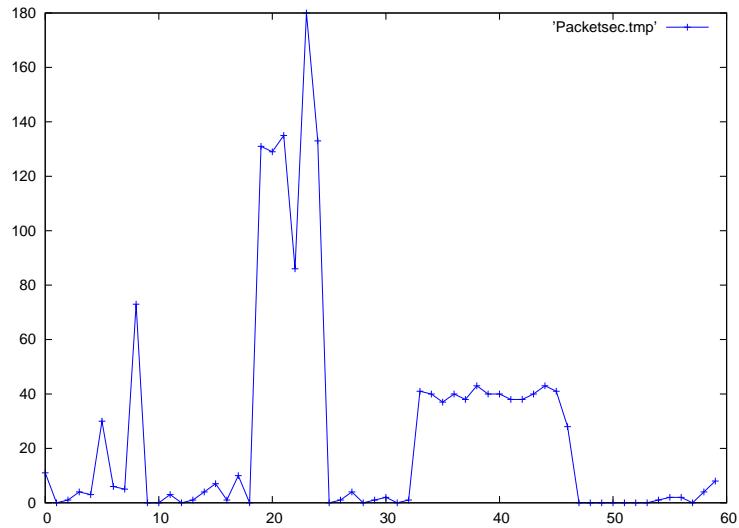


Figure H.1: Plot of Packets per seconds after the minute values for one hour of sample data.

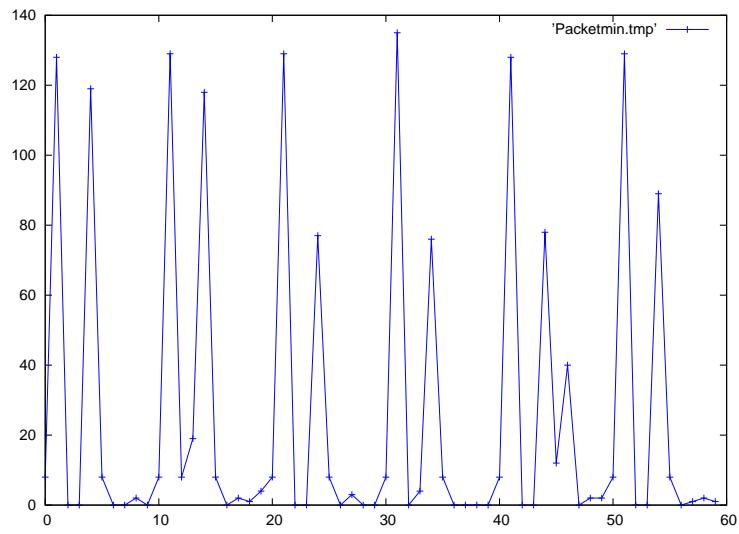


Figure H.2: Plot of Packets per minutes after the hour values for one hour of sample data.

Packets per hours after midnight Local

Our plot of packets per hours after midnight local, as seen in Figure H.4, is just as boring as the hours after midnight UTC, for the same reason. This measurement is more useful for comparing internal (outgoing) traffic, involving people or processes on the internal network contacting servers external to the network.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of packets between the two traces.

Packets per day of the week UTC

In Figure H.5, we have yet another boring plot: packets per day of the week, UTC. This is a measure that will only be interesting if comparing multiple days of data. We expect to see the same types of general patterns as we would see with the hourly measurements, on a more coarse scale: workplaces will see more activity on Monday through Friday (perhaps with noticeably less on Friday), and residences will see more activity on the weekends. The decision to base the day of the week on UTC rather than local time is somewhat arbitrary. Since the measurement is so coarse, it probably will not matter, especially if only looking at two different networks in the same country or region (the United States, or Europe, for example). If comparing networks on opposite sides of the world (one in the US to one in India, for example), then one should evaluate if using day of the week, based on local time, would be a better measurement.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of packets between the two traces.

Bytes transferred per seconds after the minute

Whereas looking at the connections per seconds after the minute provided a coarser look than the packets per seconds after the minute, the bytes transferred per seconds after the minute, as seen in Figure H.6, provides a finer grained look at the network activity. In this plot we can see that the burst of packets between 18 and 25 seconds that accounted for three to four times the number of packets than the burst from 32 to 47 seconds in Figure H.1 accounts for upwards of an order of magnitude more bytes transferred. This level of granularity is useful for differentiating interactive activity (such as telnet or SSH), which has lots of little packets, from bulk transfer activity (such as software distribution), which has lots of large packets.

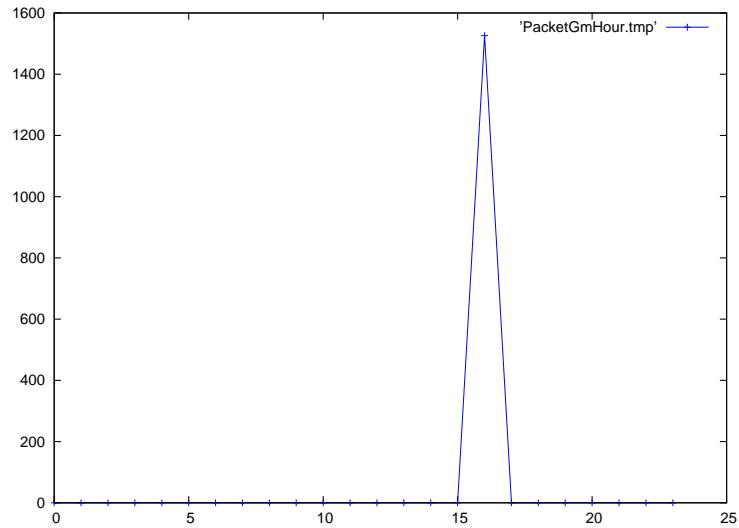


Figure H.3: Plot of Packets per hours after midnight UTC values for one hour of sample data.

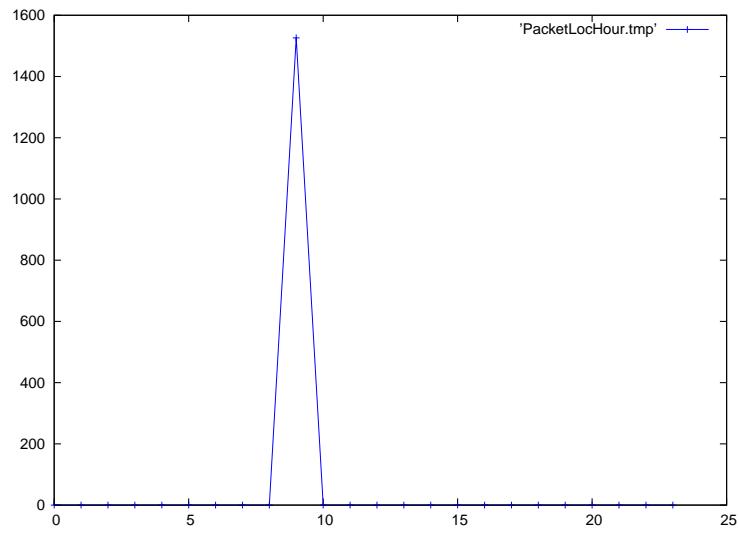


Figure H.4: Plot of Packets per hours after midnight Local values for one hour of sample data.

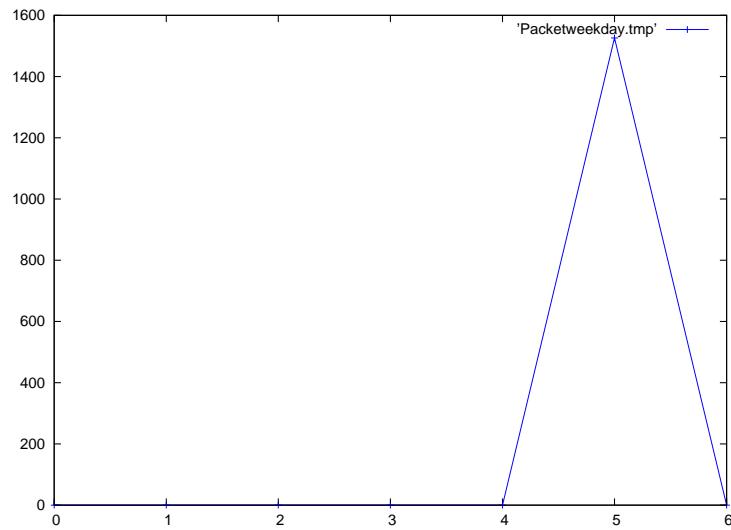


Figure H.5: Plot of Packets per day of the week UTC values for one hour of sample data.

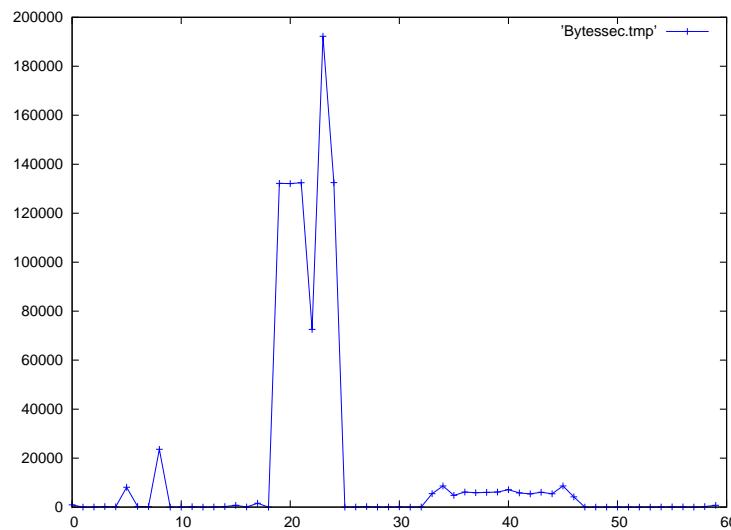


Figure H.6: Plot of Bytes transferred per seconds after the minute values for one hour of sample data.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of bytes between the two traces.

Bytes transferred per minutes after the hour

Figure H.7 shows the bytes transferred per minutes after the hour. It is analogous to packets or connections per minutes after the hour, in the same way that bytes transferred per seconds after the minute is to packets or connections per seconds after the minute.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of bytes between the two traces.

Bytes transferred per hours after midnight UTC

Figure H.8 shows the bytes transferred per hours after midnight, UTC. It is analogous to packets or connections per hours after midnight, UTC, in the same way that bytes transferred per seconds after the minute is to packets or connections per seconds after the minute.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of bytes between the two traces.

Bytes transferred per hours after midnight Local

Figure H.9 shows the bytes transferred per hours after midnight, local. It is analogous to packets or connections per hours after midnight, local, in the same way that bytes transferred per seconds after the minute is to packets or connections per seconds after the minute.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of bytes between the two traces.

Bytes transferred per day of the week UTC

Figure H.10 shows the bytes transferred per day of the week, UTC. It is analogous to packets or connections per day of the week, UTC, in the same way that bytes transferred per

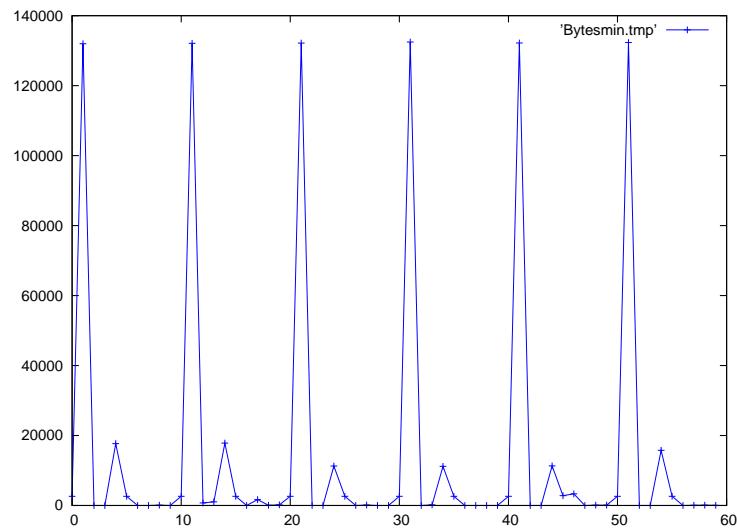


Figure H.7: Plot of Bytes transferred per minutes after the hour values for one hour of sample data.

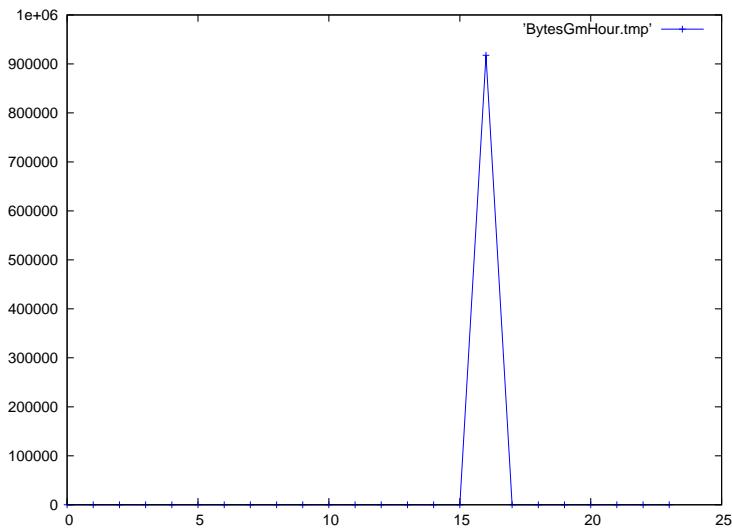


Figure H.8: Plot of Bytes transferred per hours after midnight UTC values for one hour of sample data.

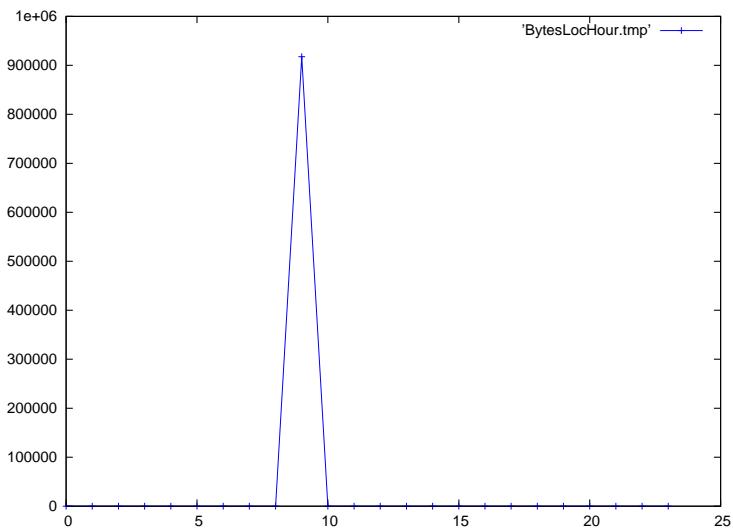


Figure H.9: Plot of Bytes transferred per hours after midnight Local values for one hour of sample data.

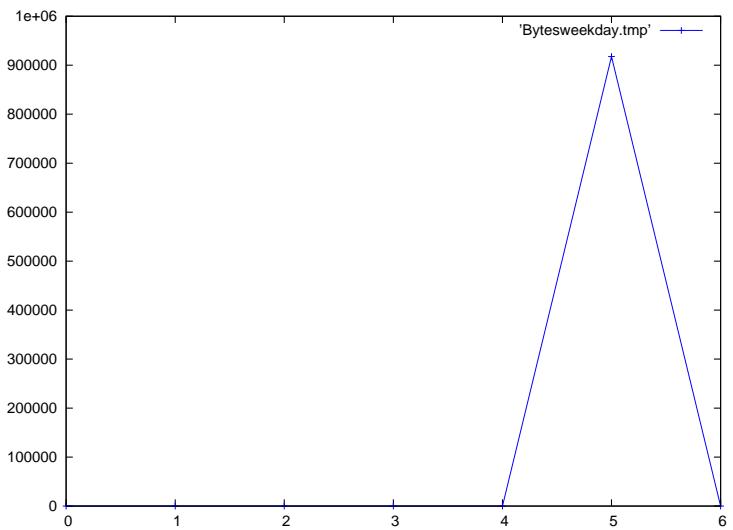


Figure H.10: Plot of Bytes transferred per day of the week UTC values for one hour of sample data.

seconds after the minute is to packets or connections per seconds after the minute.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of bytes between the two traces.

Packet sizes

The packet sizes (in bytes) are plotted out in Figure H.11. One can see that there is a handful of packets of the minimum size for TCP traffic (40 bytes), followed by a small peak at 52 bytes, which were apparently caused by a propensity of one service to use that length, probably for control information. We then see a few packets of each length used until we approach 128 bytes. Here we see a nice curve, which is likely caused by various links having a MTU around 128 bytes. After that we only have one or two packets of various lengths up to another smaller peak near 600; this is also likely due to a negotiated MTU for some connection. The odd length connections continue until we suddenly shoot up at 1500, which is the MTU for the Ethernet interface that connects this network to the Internet; this is a very common MTU for Ethernet and frame-relay devices, and hence it likely serves as the MTU for many connections, explaining the dominant spike in the graph.

As with the TTL plot, we expect that the best way to compare this is doing a discrete comparison on the weighted averages.

H.1.2 Per packet, time rates

Looking at the characteristics of individual packets in isolation gives us some basic characteristics of the network, but it does not tell us much about the behavior of the network. To start to get an idea of the behavior of the network, we look at the characteristics of packets over the last w seconds. These characteristics tell us things like how bursty our traffic is, and how diverse the traffic is over a short time period. These characteristics are influenced strongly by our prior research in the network intrusion detection field (Brugger 2007a). We calculate these characteristics for each packet that we observe, and the calculations include the current packet, and the connection that the current packet is part of (if applicable). Most of the measures are represented in rates – by taking a measured quantity and dividing it by another – in order to normalize the data for comparing across networks with different quantities of traffic.

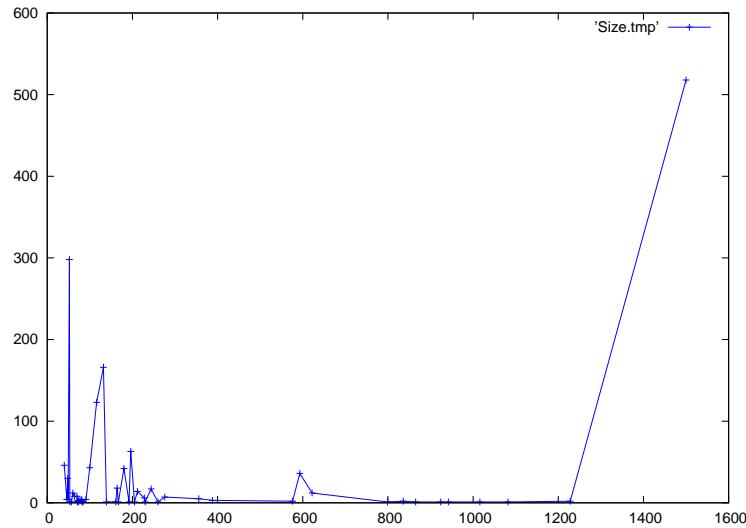


Figure H.11: Plot of Packet sizes values for one hour of sample data.

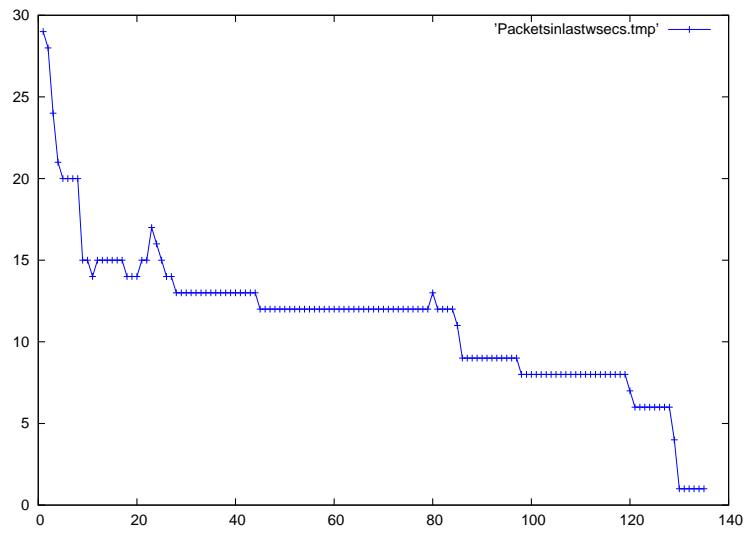


Figure H.12: Plot of number of packets over the past w seconds for one hour of sample data.

Packet rate time window

The number of packets over the past w seconds (for each packet, including this packet), shown in Figure H.12, gives us a good overview of how bursty the traffic is on this network. In our one hour of sample data, we see that many packets (nearly 30) were lonely packets in their time windows. We see a general drop-off in the number of times a given count within a window was observed, with only a couple minor increases at 23 packets in the past 30 seconds, and an even smaller spike at 80 packets in the past 30 seconds. More interestingly, we have a number of plateaus, for instance at 8 or 12 observations of a number of values. This is explained by imagining that a given connection will have a lot of traffic in a short period. Since these observations are per-packet, the first packet in the connection may be the only packet in the past 30 seconds. The second packet will make up two packets in the time window, and so forth. So the eight observations of 98 packets in the past 30 seconds are likely the same eight observations of 119 packets in the time window. They are also likely eight of the 12 observations that saw upwards of 84 packets. In essence, the plateau structure we see here is a result of the low traffic rates of this network and individual connections appearing very distinctly in the data. We expect to see a much “curvier” plot on networks with significantly more traffic.

As we look at the plot, it occurs to us that this data may be better plotted out in a cumulative distribution format (CDF). At the same time, the current format is quantitative, and it appears that it will be sufficient to do a discrete comparison on the weighted averages. Changing the format that we represent the data in would require developing a new comparison method, which we are not opposed to; however, we are inclined to avoid additional work unless there is some indication that it is necessary. We mention the different potential representation here as a suggestion for future work.

Thinking about the CDF representation does, however, remind us of the importance of scaling the data (along the y-axis) prior to comparison such that the traffic rates do not get measured with multiple metrics. Hence, when comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of packets between the two traces.

Privileged rate time window

Given the number of packets we see over a given time window, we can start to look at characteristics of those packets in that time window. In Figure H.13 we see the rate of privileged packets over the past w seconds. We calculate the privileged packet rate as the number of privileged packets in the time window divided by the total number of packets in the time window. A value of 0 means that none of the packets in the past w seconds were privileged, and a value of

1 (the maximum possible value) means that all of the packets in past w seconds were privileged.

This value is interesting as it gives an idea of proportion of traffic that goes to or from standard services (which typically run on privileged ports). For a network with a significant amount of traffic, we expect this rate will be fairly constant. Since our sample traffic is fairly sparse, we can see a range of different activities taking place on this network. Indeed, over half the traffic occurred during time windows with no privileged packets, and the remaining traffic composes a nice “S” curve, with a fairly broad middle made up of an almost even amount of privileged and unprivileged packets, and very few observations where all the packets were privileged.

The rates are plotted out in increasing order, since we are not concerned with how the rate differs from one observation to the next (although this may be an interesting thing to investigate in future work). In this form we should be able to use the same comparison technique as we did with interpacket arrival time – scaling the x-axis of both graphs to the same scale and finding the mean similarity between each point on the smaller graph to a (possibly projected) point on the larger graph.

Unprivileged rate time window

Besides the privileged packet rate over our w second window, we are also interested in the unprivileged packet rate over the time window. The definition for this is the same: the number of unprivileged packets over the past w seconds, divided by the total number of packets over the same time window. The range of possible values is also [0..1] We make this observation for each packet, and when we sort the observations and plot them out, we get Figure H.14. This figure looks to be the inverse of Figure H.13, and in this case it is; however, this is not by definition. Given our definitions in section 7.2.1, we could have packets that are neither privileged nor unprivileged which will affect the total count. In this case, since we do not have any of those, the plots end up being the inverse of each other.

As with the privileged packet rate time window, we believe that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Connection versus packet time rate window

We are also interested in the ratio between the number of connections and the number of packets. These ratios are plotted in sorted order in Figure H.15. As the ratio approaches one, it means that there are fewer packets per connection, and a ratio of one means that there is only one packet per connection. This is typically indicative of malicious activity, such as a denial of

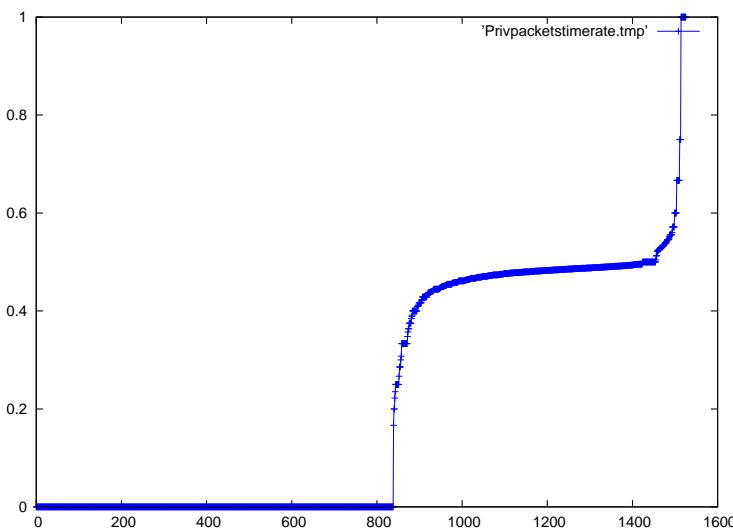


Figure H.13: Plot of number of packets to privileged services versus the number of packets over the past w seconds for one hour of sample data.

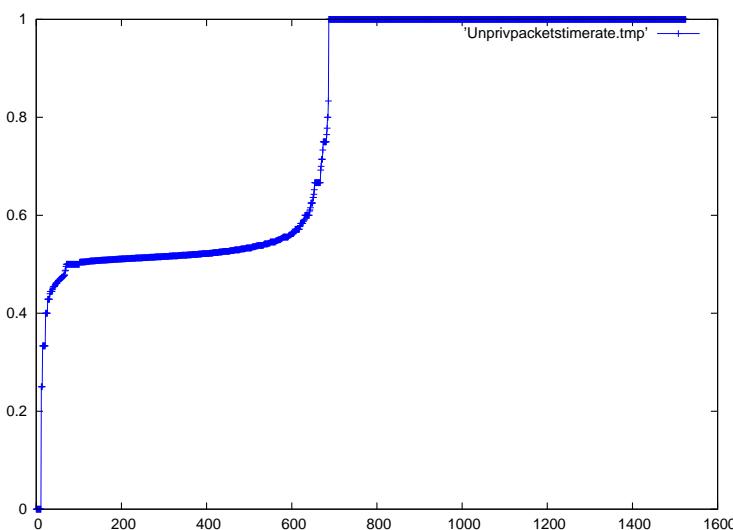


Figure H.14: Plot of number of packets to unprivileged services versus the number of packets over the past w seconds for one hour of sample data.

service, where SYN packets flood the network, or probing activity that is not responded to. It is not necessarily malicious though, as heavy UDP traffic may appear this way as well, should each packet be treated as a single connection. There must always be at least one connection making up the numerator of the connection, so the value can approach, but will never reach 0. In summary, the range of possible values is (0..1].

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of connections between the two traces.

Privileged connection versus connection time rate window

Figure H.16 shows the ratio between privileged connections and all connections, over the past w seconds. The range of possible values is zero (if none of the connections are privileged) to one (if all of the connections are). The ratios are plotted out in sorted order. As one can see, most of the time windows consisted of entirely privileged or unprivileged connections, with only a couple smaller areas of intermediate values, predominantly 0.5. We find it interesting that there is such a paucity of values; we would have expected to see more curves, particularly at the ends of the plateaus. This intuition comes from information such as the privileged packet rate, which forms a nice S-curve. We thought that the behavior could be explained by something like ICMP responses (such as “service not available”) to privileged traffic (such as DNS requests). Unfortunately, as we have already seen, this trace is (unusually) devoid of ICMP traffic (or any traffic that is neither privileged nor unprivileged), so this required further investigation, where upon we found that it came from tunneled traffic, which came in on a privileged port (SSH) and left going to a proxy running on a unprivileged port.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Unprivileged connection versus connection time rate window

The flip-side of the previous one is the ratio between unprivileged connections versus all connections, over the past w seconds, as shown in Figure H.17. Again, the possible range of values is zero (if none of the connections are unprivileged) to one (if they all are). The ratios are plotted out in sorted order. As with the privileged packet rate, above, plotting out these values in sorted order results in the inverse of Figure H.16, which is not by definition, but rather

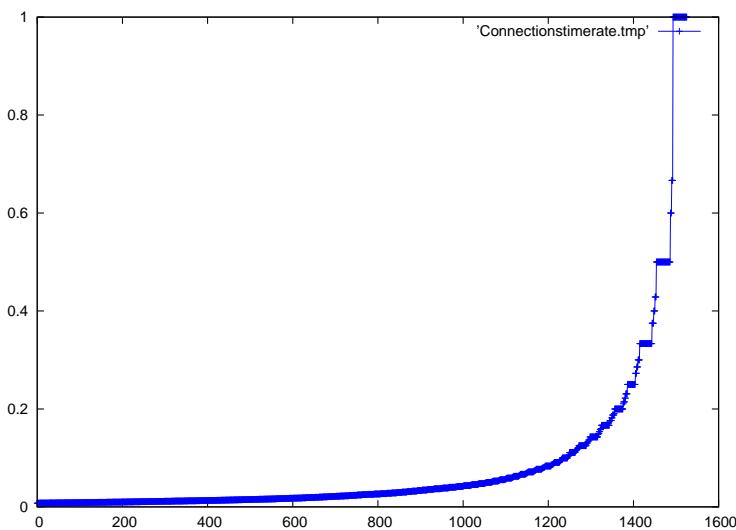


Figure H.15: Plot of ratio between connections versus packets over the past w seconds for one hour of sample data.

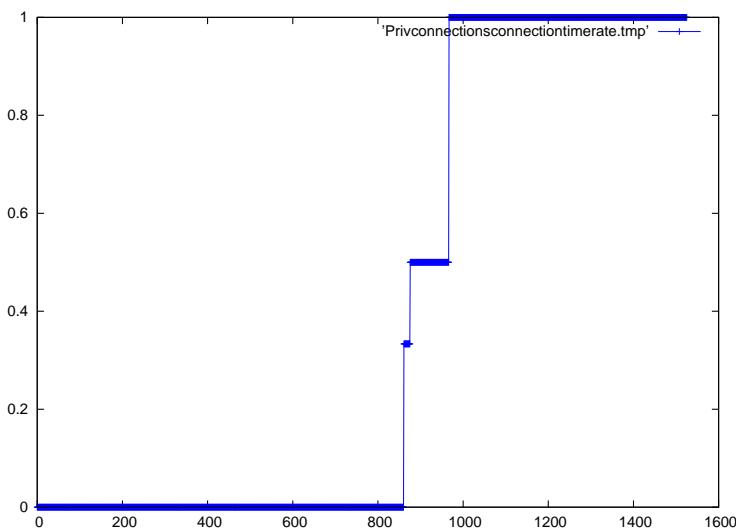


Figure H.16: Plot of ratio of privileged connections to all connections over the past w seconds for one hour of sample data.

because we do not have any connections that are neither privileged nor unprivileged.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Privileged connection versus privileged packet time rate window

Here we look at the packet rate to privileged connections, which is just the number of privileged connections versus the number of packets to those services over the past w seconds. These ratios are plotted out in sorted order in Figure H.18. This ratio will be zero when there are not any privileged connections over the past w seconds, which is something of a special case as it will be one when there was only one packet for every privileged connection, which continues to decrease as the number of packets to the same number of privileged services increases. While setting up the ratio in this manner may seem counter-intuitive (as a human analyst would likely be interested in the inverse), we do it this way to maintain a constant range of [0..1]. Since we are doing this ratio consistently, we can still achieve the same quantitative comparison between two measurements, without having to worry about adjusting the measurements for disparate traffic levels.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Unprivileged packet versus unprivileged connection time rate window

In addition to the packets to privileged services, we look at the packets to unprivileged services, as shown in Figure H.19. This plot does not look like the inverse of Figure H.18 because many of the observations (of the past w seconds) contained both privileged and unprivileged traffic. Once again, the ratio ranges from zero if there are no unprivileged connections, to one for a 1 : 1 connection versus packet ratio for unprivileged connections.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Connection versus SYN time rate window

The rate of SYN flags tells us a lot about the status of the network, particularly with regards to availability and reliability. In normal network activity we expect to see two SYNs for every TCP connection (the original SYN and the SYN-ACK). In a network denial of service

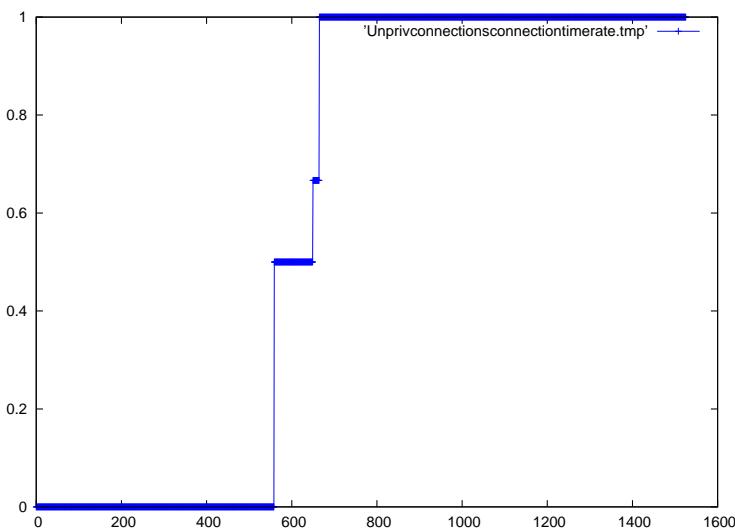


Figure H.17: Plot of ratio of unprivileged connections to all connections over the past w seconds for one hour of sample data.

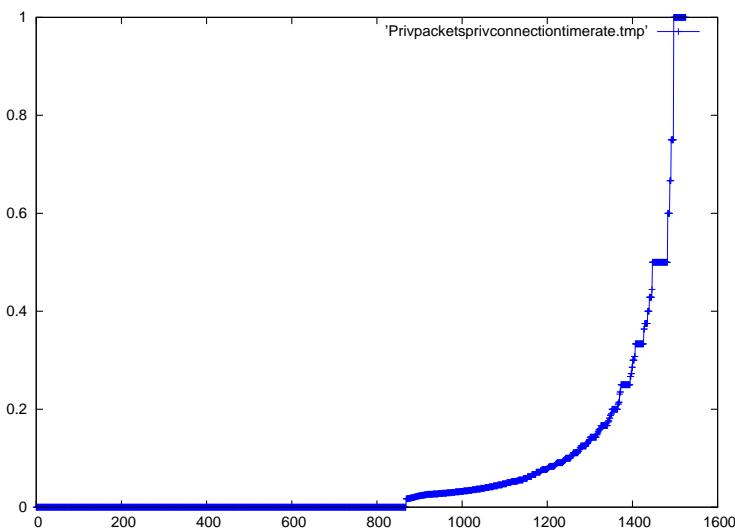


Figure H.18: Plot of ratio of privileged connections to privileged packets over the past w seconds for one hour of sample data.

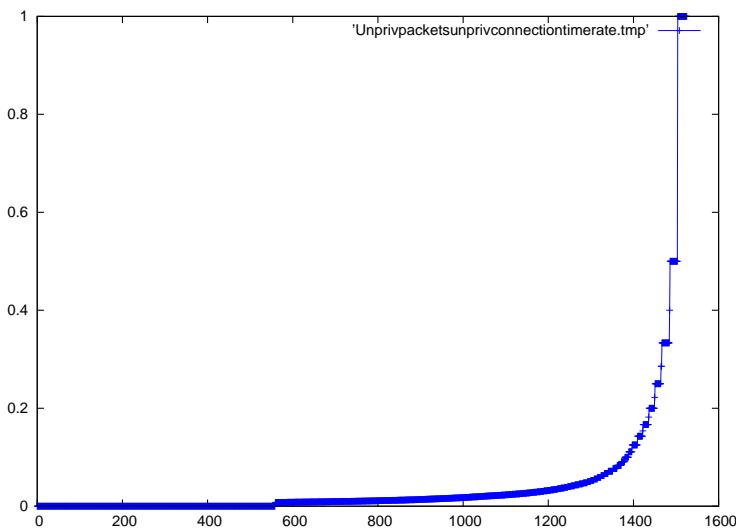


Figure H.19: Plot of ratio of unprivileged connections to unprivileged packets over the past w seconds for one hour of sample data.

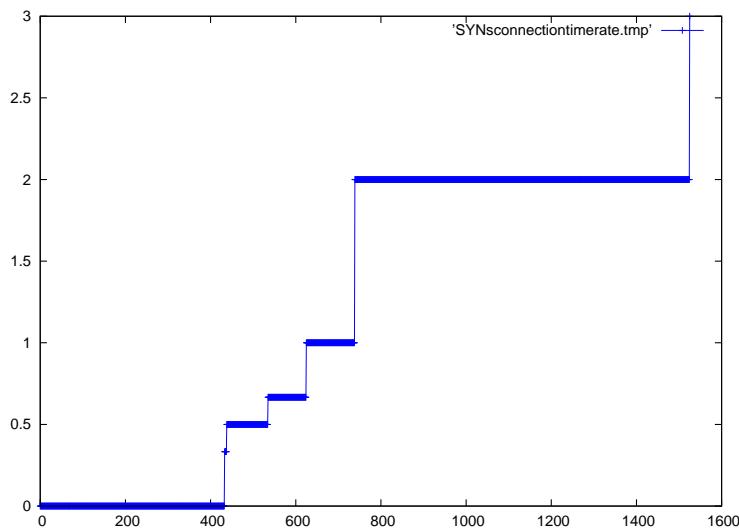


Figure H.20: Plot of SYN flags versus connections over the past w seconds for one hour of sample data.

attack, during probing activity, or in the event of blocked connections, the initial SYNs will not be responded to, resulting in a ratio closer to one. On the other hand, in a flash crowd scenario, clients may resend their SYNs multiple times, resulting in a ratio above two. The plot of these ratios on our sample data can be seen in Figure H.20.

Since we are only looking measuring SYNs over the past 30 seconds, and connections can (and do) last longer than 30 seconds, we will also see a number of ratios in the range [0..2] for normal connections, which just represents the inclusion of these longer running connections in our ratio. Unfortunately, when looking at an individual ratio of say, one, we can not tell if this was due to a denial of service attack, or one normal connection being established along with one normal connection whose SYNs had passed beyond the w second window. While this is true in isolation, taking the ratios as a complete curve is much more telling, as the denial of service attack will likely show up in a long line of ones, whereas a measurement from a couple normal connections will likely look more like part of a curve between zero and two.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Connection versus RST time rate window

In a perfect world, connections would never need to be reset (RST). Such as it is though, the TCP protocol allows that problems will arise and connections may be shut down in an irregular manner by sending a packet with a RST flag. As such, the rate of RSTs on a given network traditionally provides an idea of how reliable it is. RSTs may also be used in a variety of attacks, or in response to a variety of attacks. We look at the number of RST flags versus the number of active connections over the past w seconds. In normal operation, such ratios will only span between zero (if there are no RSTs) to one (if every connection has an RST). A connection could have multiple RSTs, for instance if both sides send a RST simultaneously, due to TCP implementation problems (such as the RST being ACKed), or due to malicious activity. While RFC793 explicitly states that an RST closes a connection immediately (Postal 1981), our network connection metric code (see chapter 6) will treat a stream of RSTs as part of a single connection until such time as a new SYN is sent with the same host and port pairs.

The plot of the RST to connection ratios for our one hour of sample data can be seen in Figure H.21. This is a very boring plot, considering that we only had one ratio of 0.5 and three measurements at 1, owing to a very low rate of RSTs in our sample data. Nevertheless, we believe that we can treat this as we do our other sorted ratio values, by scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph

will provide a good similarity metric.

Connection versus FIN time rate window

As with SYN and RST flags, we look at the ratio of FIN flags to the number of active connections over the past w seconds. Much like SYN flags, on well behaved connections, this ratio will be 2 (one from each side of the connection). We will have a number of time windows where there are (potentially many) more active connections than closing connections ,resulting in ratios over the time window that are below 2. This behavior can be seen in our sample data in Figure H.22. Values above 2 indicate a problem, which could come from a bad TCP implementation, network unreliability (FINs end up getting resent), or malice.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Connection versus PSH time rate window

The PSH flag is significantly different from the other TCP flags we look at (we are not looking at the ACK flag), in that it is typically used many times per connection. As such, we flip the ratio and look at the number of active connections versus the number of PSH flags over the past w seconds. Our plot of these values is shown in Figure H.23. As seen, the majority of these ratios are indeed below 1, with a notable plateau at 1, and only a few time windows where the ratio exceeded 1. In the event there are no PSH flags in a given time window, we set the ratio to zero.

Given our security background, we generally do not find PSHs too interesting. They are, however, important as they capture some of the semantics of the network: how often one of the endpoints thought what it had was important enough to push through right away. As such, they seem important for comparing how similar two network traces are.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Connection versus establishment error time rate window

Connection errors are any errors that occur during the TCP three-way handshake. While these are commonly of interest when networks get overloaded – either due to denial of service attacks or flash crowds – they are a typical part of contemporary networks, as blocked connections due to firewalling may also appear as establishment errors. We look at the rate of

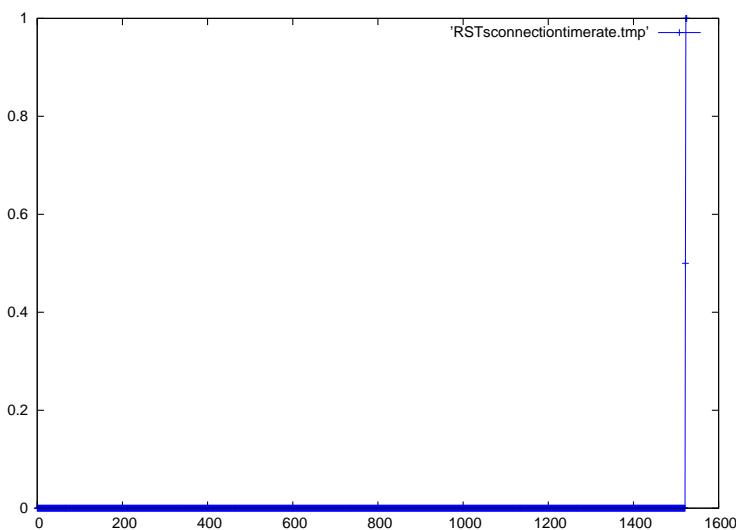


Figure H.21: Plot of RST flags versus connections over the past w seconds for one hour of sample data.

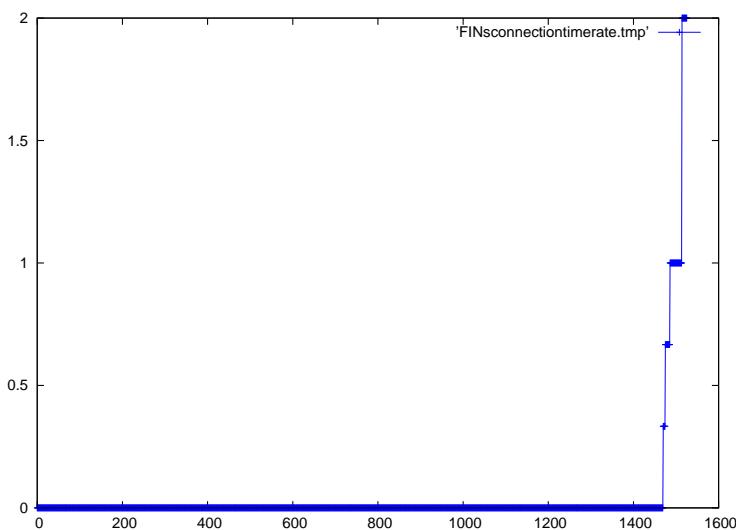


Figure H.22: Plot of FIN flags versus number of connections over the past w seconds for one hour of sample data.

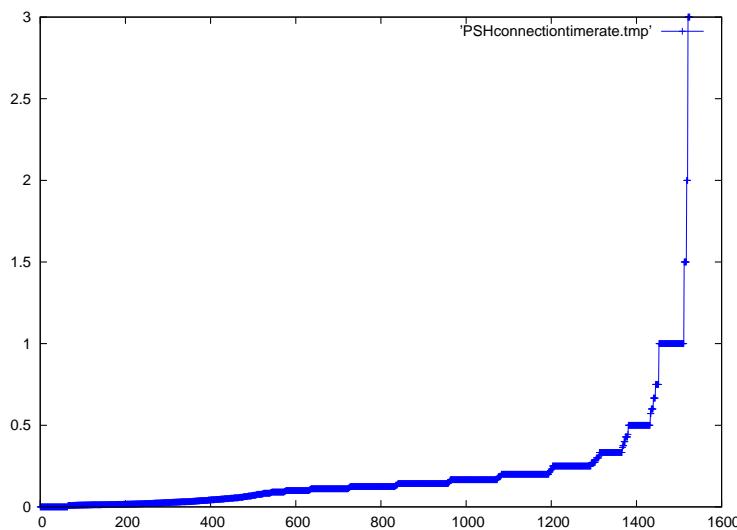


Figure H.23: Plot of number of connections versus PSH flags over the past w seconds for one hour of sample data.

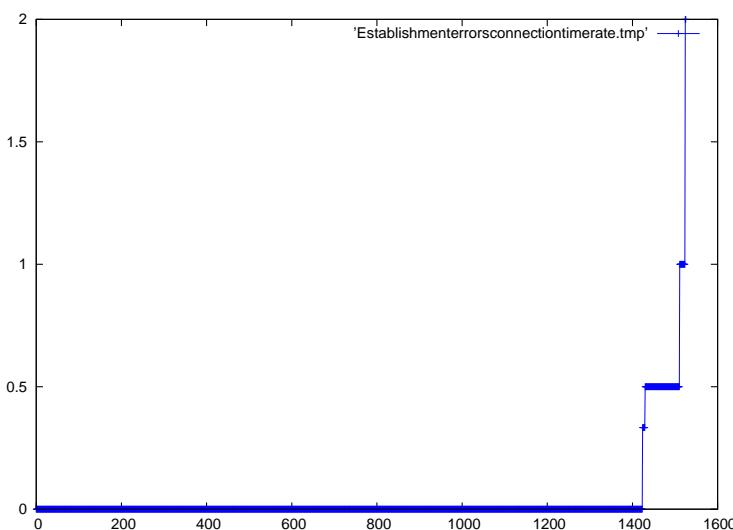


Figure H.24: Plot of establishment errors versus active connections over the past w seconds for one hour of sample data.

establishment errors versus active connections over the past w seconds. The plot of these ratios on our sample data can be seen in Figure H.24.

As expected, the observed value over most of our time windows is zero, with a few values approaching or at one, and a couple at two from attacks that were blocked and re-tried. Theoretically, there is no upper limit to how high this ratio could go; in practice, we expect that our sample data is generally representative of what most networks see. While the plot of these ratios alone does not tell us what is happening on the network, we anticipate that they will be insightful when comparing how similar two network traces are.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Connection versus other error time rate window

Other errors are fairly rare: they consist of TCP protocol violations such as setting the SYN flag, or not setting the ACK flag in an established connection. Like establishment errors, we look at their ratio with active connections over the past w seconds; and, like establishment errors, this ratio can range from 0 to inf, however we expect that it will typically be very low. As can be seen in the plot of our sample data – Figure H.25 – we did not experience any.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Connection versus disconnection error time rate window

Disconnection errors are any TCP protocol errors that occur when the connection is being torn down (once the first FIN is sent). Disconnection errors may be slightly more common than “other errors”, however, as can be seen in Figure H.26, we still did not have any in our sample data. Still, we expect that their ratios to active connections, which may span from 0 to inf over the past w seconds will be a useful way to represent them for comparison when they do occur.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

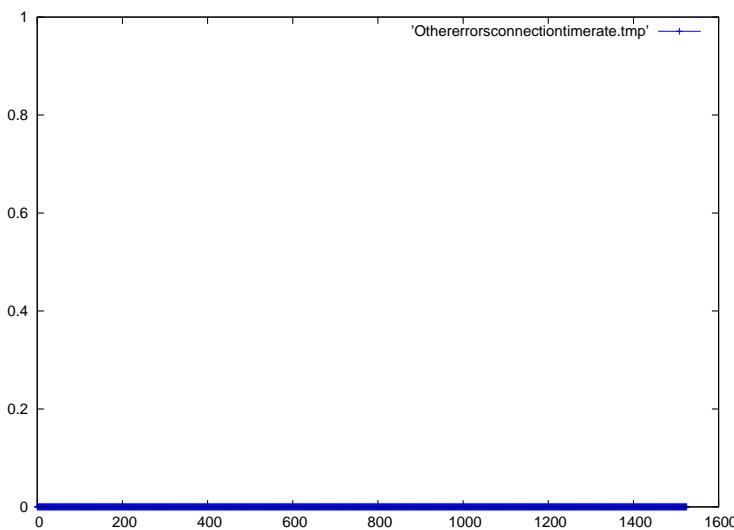


Figure H.25: Plot of other errors versus active connections over the past w seconds for one hour of sample data.

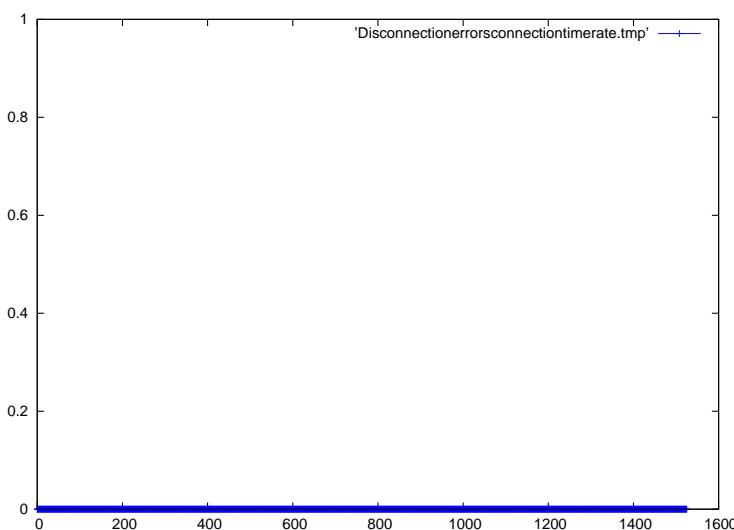


Figure H.26: Plot of disconnection errors versus active connections over the past w seconds for one hour of sample data.

Average duration time rate window

Figure H.27 shows the average duration, in seconds, for connections active over the past w seconds. This is a different view of connection duration than we typically think of it as we include connections that are active, so the values do not represent the total duration of all the connections included in the average. Furthermore, it is a per-packet measurement, not a measure made at the conclusion of the connection. While the initial measurement made in an observation run, and the first measurement after w seconds of no network activity, will be zero, most of the measurements will be non-zero. Since we only calculate connection duration for those connections that we observe from inception, the largest value possible is equal to our length of our observation run, however practically (as is the case here), it will be much less.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

H.1.3 Per packet, packet rates

In the last subsection we looked at measurements that were made per packet averaged over the number of active connections in the past w seconds. Now we shift our focus slightly and look at the per packet measurements averaged over the past n packets.

Privileged rate data window

Figure H.28 shows the rates of packets to privileged services out of the past n packets for each packet. Since we are looking at what percentage of the packets are privileged, this value is naturally constrained to the range [0..1]. The stair-stepping is due to the limited number of possible values (anywhere from 0 to 50 of the past 50 packets).

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Unprivileged rate data window

Figure H.29 shows the rates of packets to unprivileged services out of the past n packets for each packet. As above, we are naturally constrained to the range [0..1], with stair-stepping due to the limited number of possible values. As with the other privileged/unprivileged plots, this one is the inverse of the privileged plot above only by virtue of there not being any traffic that is neither privileged nor unprivileged.

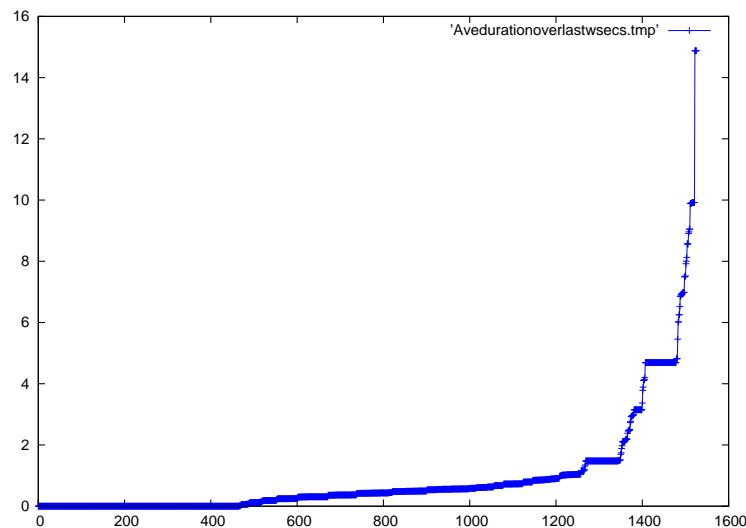


Figure H.27: Plot of average duration of connections active in the past w seconds for one hour of sample data.

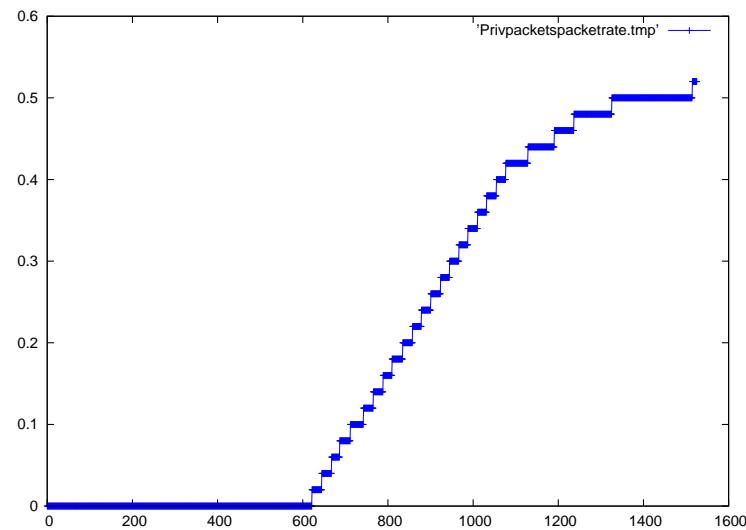


Figure H.28: Plot of rate of privileged packets out of the past n packets for one hour of sample data.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

H.1.4 Per connection initiation

Now we turn our attention to metrics that we look at per connection. We start with basic metrics that we can measure at the time a connection is initiated. It is worth noting that we output this information as soon as we see a new connection, whereas other connection metrics, as discussed in chapter 6 and demonstrated in section H.1.6, are output when we have determined a connection has closed. This means that the number of connections we see in our temporal measurements may be larger than the number of connections we see for the other connection metrics, as those metrics will not be given on the connections that are still active when collection ends. The reason we output these metrics at connection initiation time instead of waiting until we determine the connection is closed is that a lot of what we are looking at is time sensitive, and we can say with certainty when a connection is initiated, whereas it may be vague as to when exactly a connection ends.

Interarrival time of connections

As shown in Figure H.30, we have plotted out the connection interarrival times in order from smallest to largest, just as we did for packet interarrival times. The curve is obviously much less dramatic, and is indicative of the low level of use of this network, given how large some of the deltas are (significantly over five minutes). Despite this significant difference from packet interarrival times, the same technique we developed for comparing packet interarrival times should work here.

Connections per seconds after the minute

The intuition to look at connections, in addition to packets, is that connections give us insight to the activity – the people or processes that actually do something to initiate activity on the network. Looking at the packet counts tells us more about the level of activity on the network. By using both in our comparisons, we can differentiate a network that has a lot of small connections (for instance, serving up web pages via HTTP 1.0), versus one that has a few large connections (for instance, distributing software tarballs or handling VPN connections). When we look at the activity in our sample data, seen in Figure H.31, we considered that going down to this level of detail (per second) may not be informative; however, this should be ascertained when

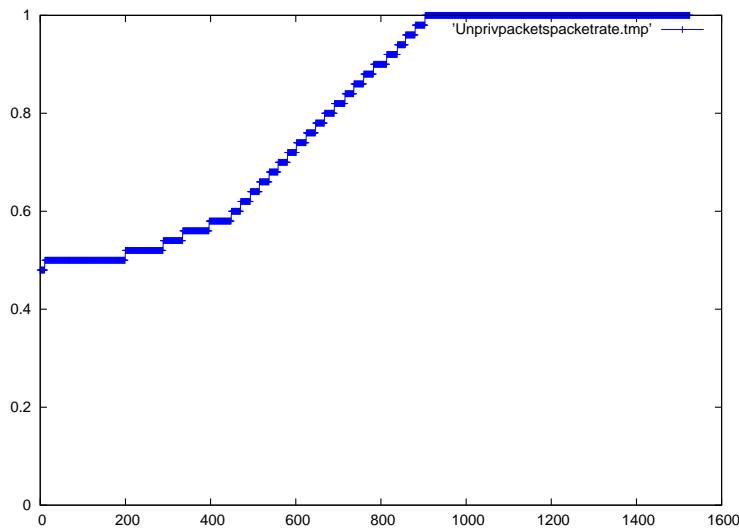


Figure H.29: Plot of rate of unprivileged packets out of the past n packets for one hour of sample data.

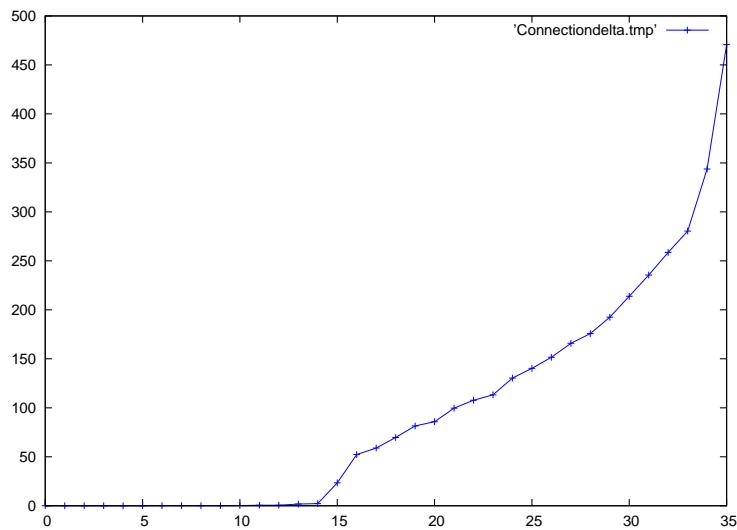


Figure H.30: Plot of Interarrival time of connections values for one hour of sample data.

we generate the weights for the metrics. We should note that the time we use for a connection is the timestamp of the first packet in the connection. This is intuitive as it is the time that action was taken on the part of a person or process to initiate the connection.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of connections between the two traces.

Connections per minutes after the hour

Figure H.32 shows the connections per minutes after the hour. It is analogous to packets per minutes after the hour in the same way that connections per seconds after the minute is to packets per seconds after the minute. In this case though, we can anticipate that the peaks seen in the figure, however minute (no pun intended), are actually indicative of patterns of activity – at least much more so than was the case with connections per seconds.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of connections between the two traces.

Connections per hours after midnight UTC

Figure H.33 shows the connections per hours after midnight, UTC. It is analogous to packets per hours after midnight, UTC, in the same way that connections per minutes after the hour is to packets per minutes after the hour, and hence has the all the data concentrated in the one hour we collected our sample from.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of connections between the two traces.

Connections per hours after midnight Local

Figure H.34 shows the connections per hours after midnight, local. It is analogous to packets per hours after midnight, local, in the same way that connections per minutes after the hour is to packets per minutes after the hour, and hence has the all the data concentrated in the one hour we collected our sample from.

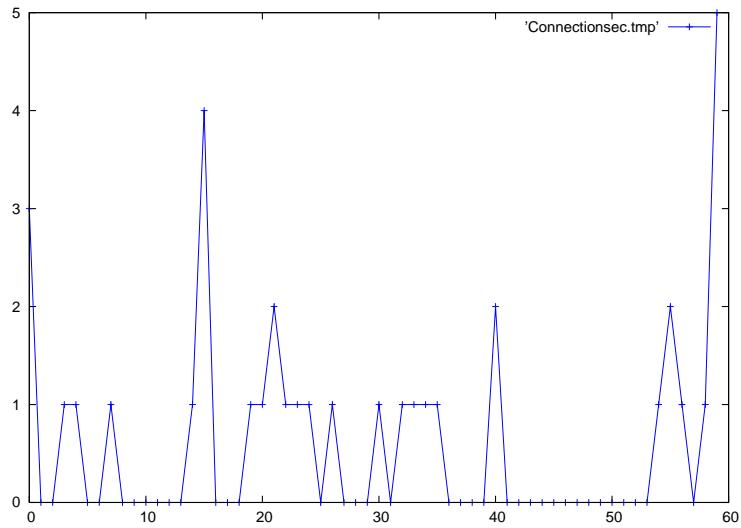


Figure H.31: Plot of Connections per seconds after the minute values for one hour of sample data.

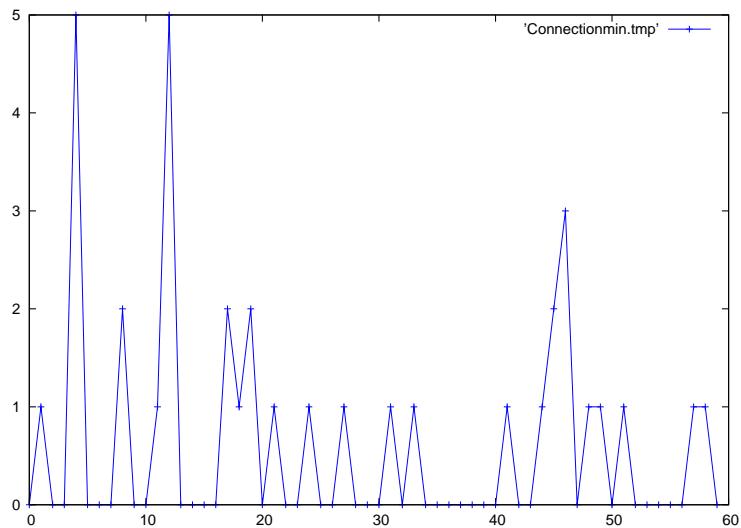


Figure H.32: Plot of Connections per minutes after the hour values for one hour of sample data.

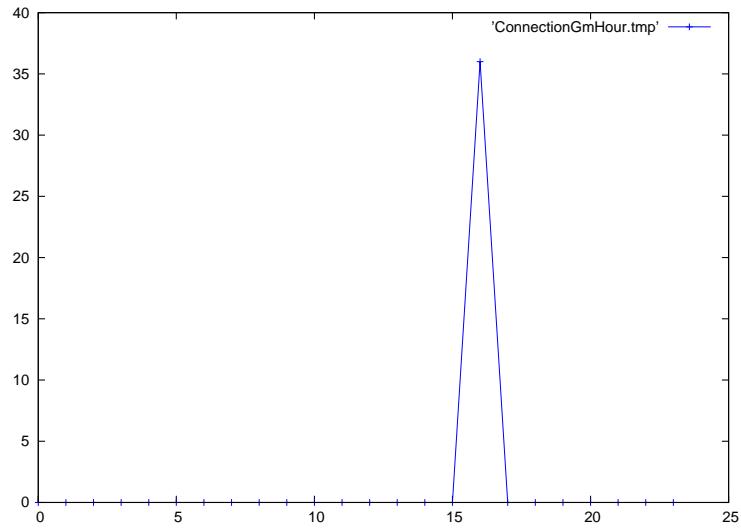


Figure H.33: Plot of Connections per hours after midnight UTC values for one hour of sample data.

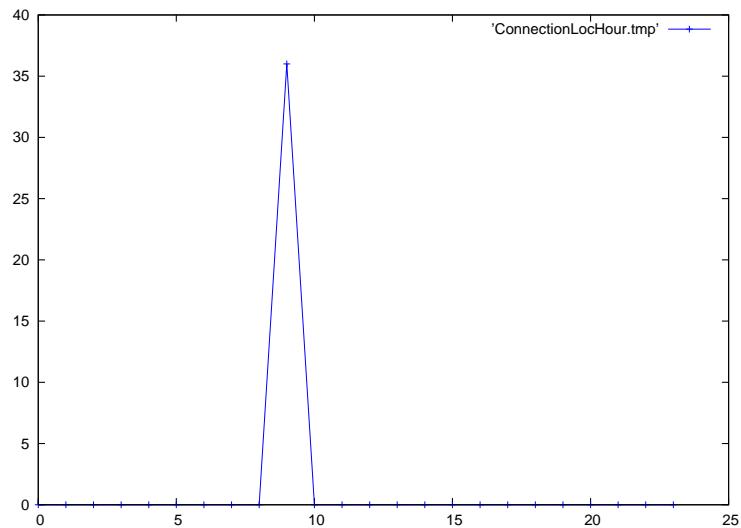


Figure H.34: Plot of Connections per hours after midnight Local values for one hour of sample data.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of connections between the two traces.

Connections per day of the week UTC

Figure H.35 shows the connections per day of the week, UTC. It is analogous to packets per day of the week, UTC, in the same way that connections per minutes after the hour is to packets per minutes after the hour, and hence has the all the data concentrated in the one day we collected our sample from.

As with our other temporal measurements, above, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of connections between the two traces.

Connection source port

While we originally looked at the source port of connections as a discrete metric, the thought occurred to us that most operating systems assign source ports as sequentially increasing across the range of possible source ports. That being the case, one would expect to see what looked like curves if we plot out the source port number as the independent variable (x-axis), and the number of connections observed that used that source port as the dependent variable (y-axis). Just such a plot was done on our one hour of sample data, and can be seen in Figure H.36. Obviously, this measurement is only made for TCP and UDP connections.

We do not have enough points in our sample data to evaluate how well this intuition will pan out, but we do notice a concentration of connection source ports in the 1024–5000 range. The spikes are due to UDP traffic where we treat each packet as a separate connection. We believe that doing a discrete comparison on the weighted averages, as we do for TTL (which, the reader may recall, we also look at both as a discrete and a continuous measurement), will provide a useful comparison metric.

When comparing this metric between traces, we will scale the domain of the second trace based on the ratio of the total number of connections between the two traces.

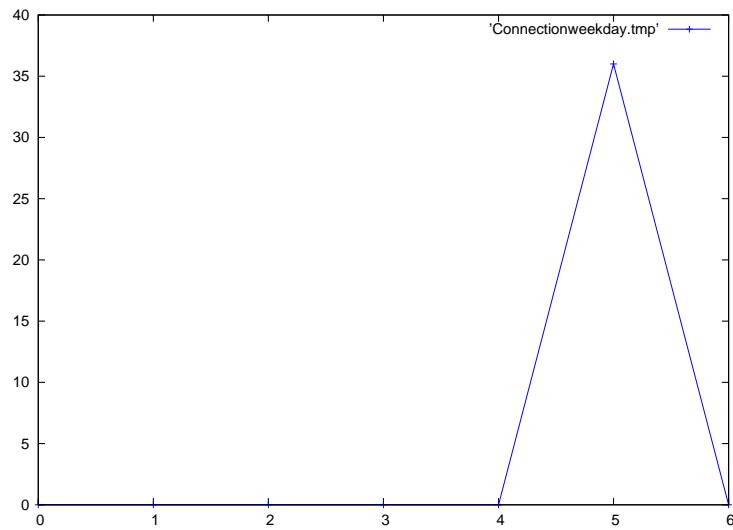


Figure H.35: Plot of Connections per day of the week UTC values for one hour of sample data.

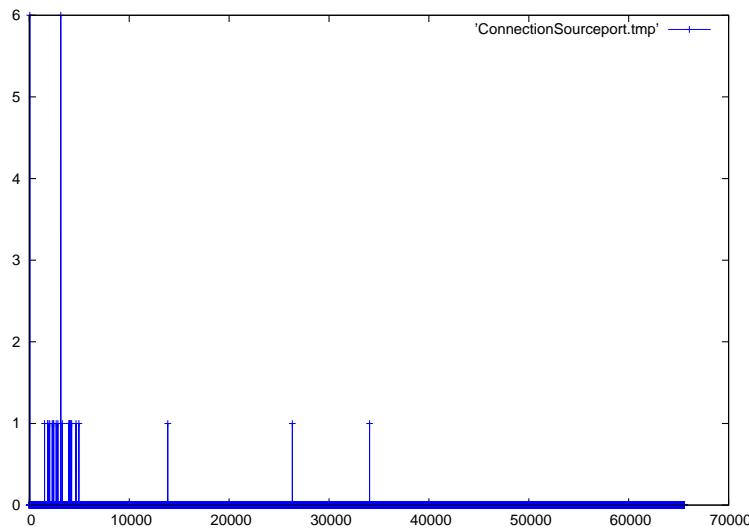


Figure H.36: Plot of connection source port values for one hour of sample data.

H.1.5 Per connection initiation, connection rates

In the last subsection we looked at the intrinsic metrics that we can measure at connection initiation. Now we turn to the connection rates over the past m connections, measured at the inception of connections. If we have not yet seen m connections, this will be the number of connections that we have seen; this is important if we are using m in calculating some other value.

We include connections that we do not have complete information on (such as connections that were already established when we began observing the connection) in the count of connections used to calculate the rates.

Connection versus packet, connection rate window

We begin by looking at the ratio of connections to packets, but instead of looking at this ratio for each connection individually, we look at its average over the past m connections. This should give us an idea of how packet intensive flows on this network are, on average. As we can see in Figure H.37, there is not a great deal in the variation of average packet counts per flow on our sample network, save for the outlier on the far right, which was actually the first observation we made, before we had enough data to establish any trends.

Since there must be at least one packet per connection, these values will be constrained to the range $(0..1]$. As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Privileged connection rate, connection rate window

We examine the privileged connection rate by taking the number of privileged connections over the past m connections and dividing it by m . The result of these observations, plotted out in sorted order, is shown in Figure H.38.

Since the number of privileged connections can not exceed the total number of connections, these values will be constrained to the range $[0..1]$. As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Unprivileged connection rate, connection rate window

We examine the unprivileged connection rate by taking the number of unprivileged connections over the past m connections and dividing it by m . The result of these observations,

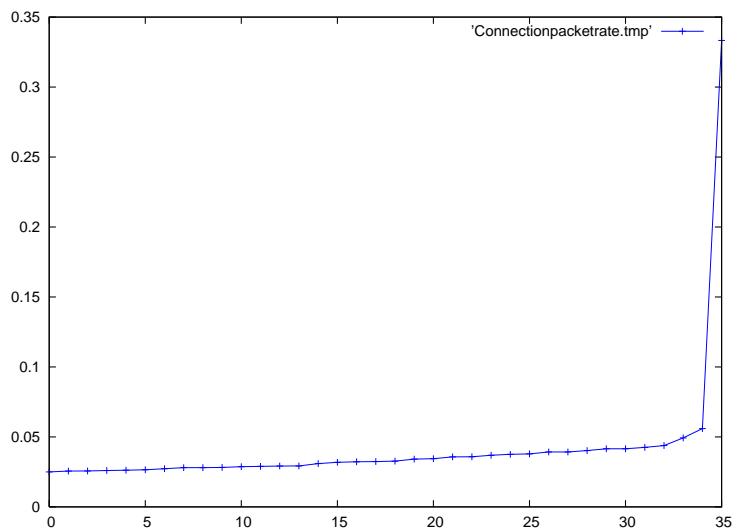


Figure H.37: Plot of connections versus number of packets, over the past m connections, for one hour of sample data.

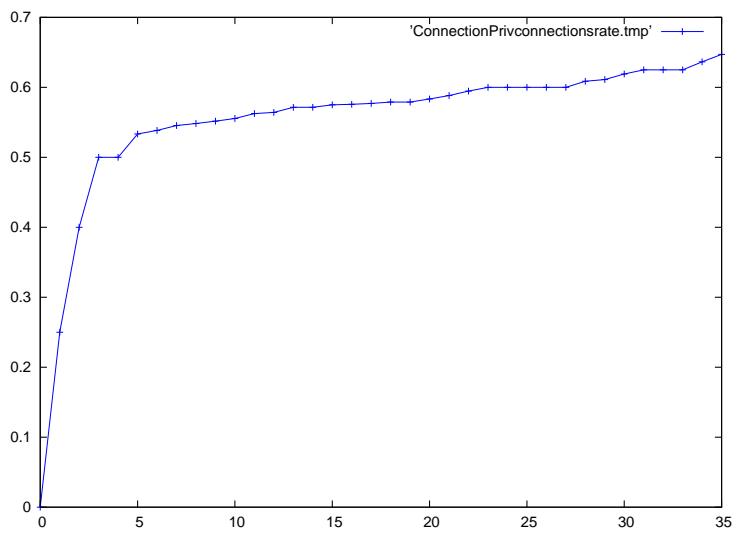


Figure H.38: Plot of privileged connections versus connections, over the past m connections, for one hour of sample data.

plotted out in sorted order, is shown in Figure H.39. As with our previous pairs of privileged and unprivileged data, this graph is the inverse of Figure H.38 only by virtue of not having any connections that are neither privileged nor unprivileged.

Since the number of unprivileged connections can not exceed the total number of connections, these values will be constrained to the range [0..1]. As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Privileged connection versus privileged packet, connection rate window

Now we turn our attention to the behavior of privileged connections specifically by looking at the number of privileged connections versus the number of packets in those connections, out of the past m connections. The results are plotted out in sorted order in Figure H.40. Note that even if a connection is unprivileged, it will still report this ratio for the prior connections that were privileged. If there are no privileged connections in the previous m , this ratio will be reported as zero (as you can see, there is one occurrence of this in the test data, which happened to be the first connection).

Since the number of packets must be greater than the number of connections, these values will be constrained to the range [0..1]. As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Unprivileged connection versus unprivileged packet, connection rate window

In a switch from previous pairs of privileged/unprivileged graphs, Figure H.41, which shows the ratio of unprivileged connections to packets in those connections, over the past m connections, looks nothing like its privileged counterpart. This is because we are looking at the given value (connections versus packets) only on the unprivileged connections, although we make the measurement for every connection. When we contrast this graph with Figure H.40, we can see that most of the packets on this network belong to unprivileged connections.

Since the number of packets must be greater than the number of connections, these values will be constrained to the range [0..1]. As before, in the absence of any unprivileged connections, the ratio is defined as zero. As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

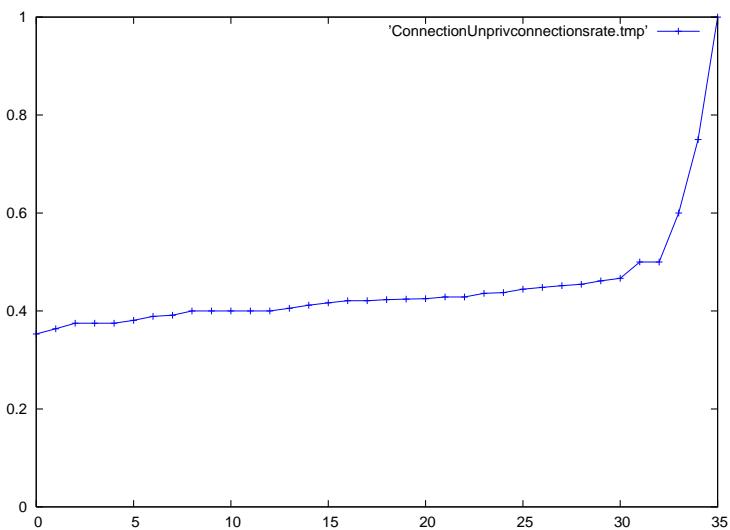


Figure H.39: Plot of unprivileged connections versus connections, over the past m connections, for one hour of sample data.

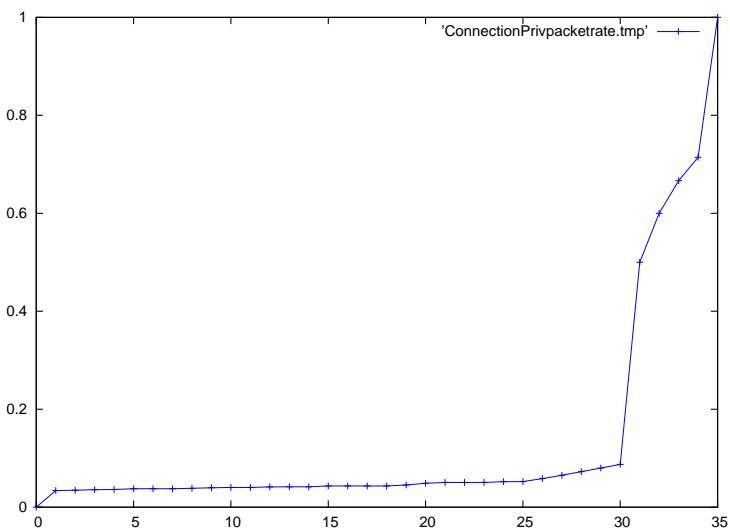


Figure H.40: Plot of privileged connections versus privileged packets, over the past m connections, for one hour of sample data.

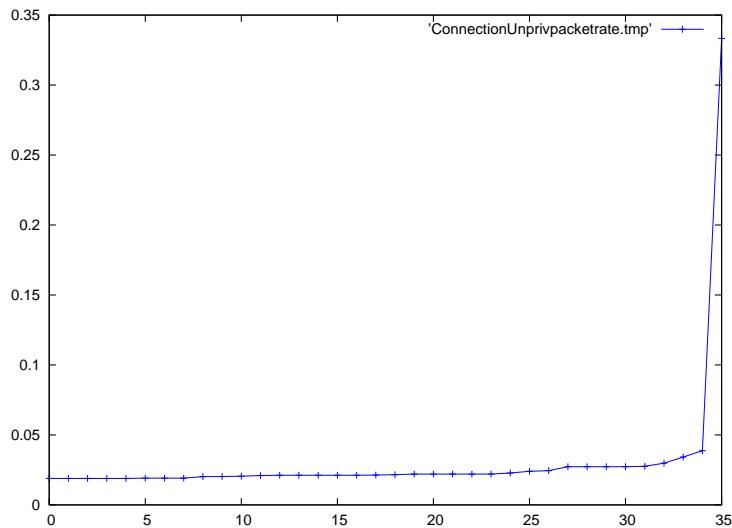


Figure H.41: Plot of unprivileged connections versus unprivileged packets, over the past m connections, for one hour of sample data.

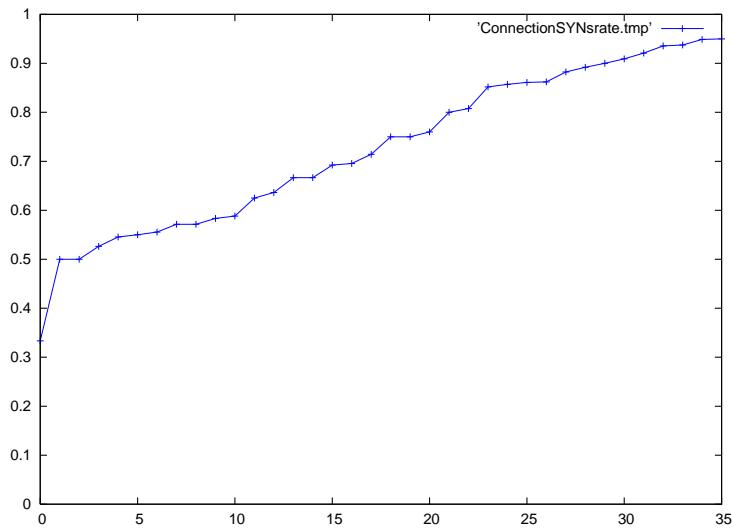


Figure H.42: Plot of SYN flags versus the number of connections, over the past m connections, for one hour of sample data.

SYN flags versus connection, connection rate window

By looking at the ratio of SYN flags to connections, we should be able to get some idea as to the health of the network. On a well behaved TCP-dominate network, we should see 2 SYNs per connection. This number may be larger in the event of connection establishment difficulties. More frequently, however, it will be lower due to non-TCP connections, or the inclusion of TCP connections that were already active when we began monitoring. We expect that such factors will average over time. We leave it as a topic for future research to see how results are affected should we measure this only with respect to TCP connections that we see from inception.

Why we think this will be informative is in a SYN-flood style denial of service attack, we expect this measure will quickly trend towards one. Conversely, in a flash-crowd situation where the server can not keep up with requests, we expect this will trend above two.

Our much more boring data is plotted out in sorted order in Figure H.42. As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

RST flags versus connection, connection rate window

TCP connections should only be reset (RST) in exceptional situations. As such, the rate of RST flags on the network gives us a good idea of the health of that network. RST flags may also be used in a variety of network attacks. In our one hour of sample data, we only saw three RSTs; this is reflected when we plot out the RST rates (number of RST flags in the past m connections, divided by m), as shown in Figure H.43; the first little hop was the first RST as observed when a different connection was initiated, and the subsequent connection (which has a slightly lower value, because the denominator was one larger); we then jump to the far right with two more RSTs, and the subsequent connections to that one (again, with diminishing values due to the growing denominator).

While technically a connection can only have a single RST, if we get subsequent RSTs with the same host and port endpoints, we group those together as their own connection. As such, there could theoretically be many more RST packets than connections, so there is no upper bound on this ratio. In practice, however, we expect this will be near zero for most networks, as we see here.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

FIN flags versus connection, connection rate window

Like SYN flags, we expect to see two FIN flags per TCP connection. To get a feel for the FIN flag rate correctness, we look at the ratio of FIN flags to connections over the connection window. This ratio will approach two if all the previous connections in the window are TCP, well behaved, and closed. It will not actually reach two in that case, since we count the connection that was just initiated as one of the m connections in the denominator. In theory, it could be higher, if one end of a connection retransmitted its FINs many times, however this would be truly exceptional. We expect a typical network will look more like the plot of values for our sample data as plotted out in sorted order in Figure H.44.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Connection versus PSH, connection rate window

Push (PSH) flags behave quite differently from the other TCP flags we look at in that we tend to see many PSH flags per TCP connection (whereas SYN, RST, and FIN should all occur a limited number of times). As a result, instead of looking at the flags versus connections, we flip the ratio around to look at the number of connections versus PSH flags. Since we will typically have more PSH flags than connections, we keep the ratio under 1. In the extremely unlikely event that there are no PSH flags, we define the value to be 0. These ratios from our hour of sample data can be seen plotted out in sorted order in Figure H.45. Push flags are used when the application wants the network buffer sent now, so the rate they appear on the observed networks should give us a partial idea about the types of applications running over this network; for example, interactive applications (like remote terminal programs) will have a very low ratio, whereas less interactive applications (like most file transfers) will cause this ratio to be higher.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Establishment error versus connection, connection rate window

Errors are something that, by definition, should not happen, but do. Establishment errors are probably the most common because if two hosts have a problem talking to each other, it is likely to be a problem from the outset. Establishment errors may or may not mean that the connection finishes getting established, and numerous errors may occur during the establishment of a connection, meaning that while this ratio will likely be close to zero most of the time, there

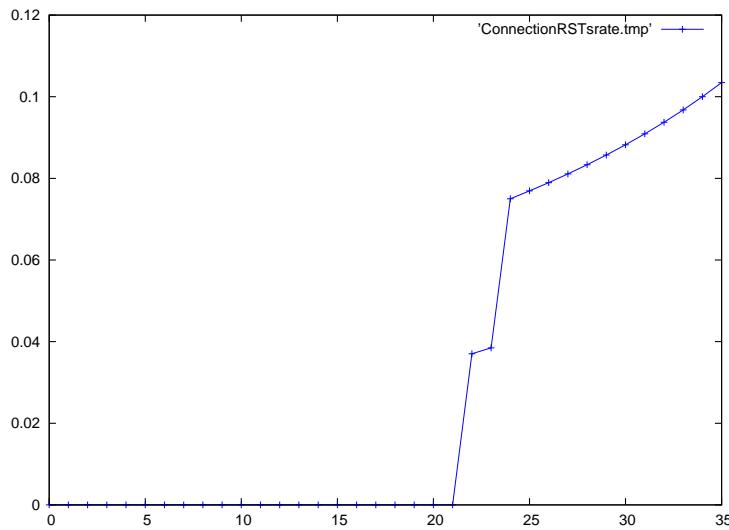


Figure H.43: Plot of RST flags versus the number of connections, over the past m connections, for one hour of sample data.

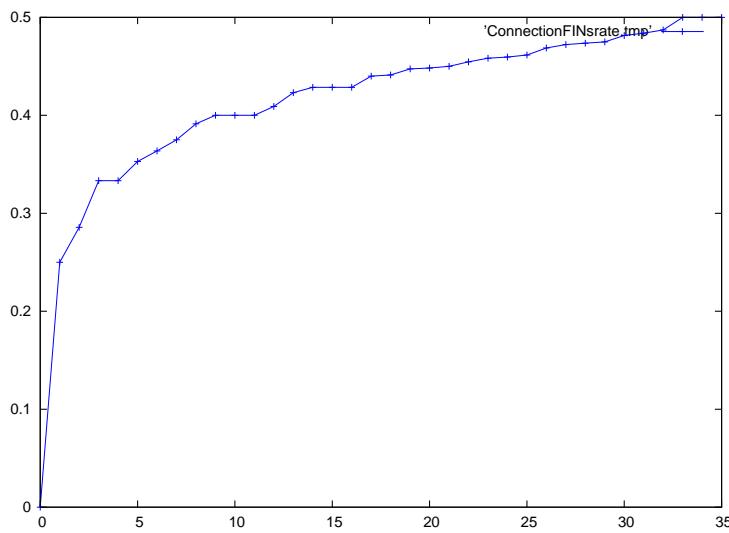


Figure H.44: Plot of FIN flags versus the number of connections, over the past m connections, for one hour of sample data.

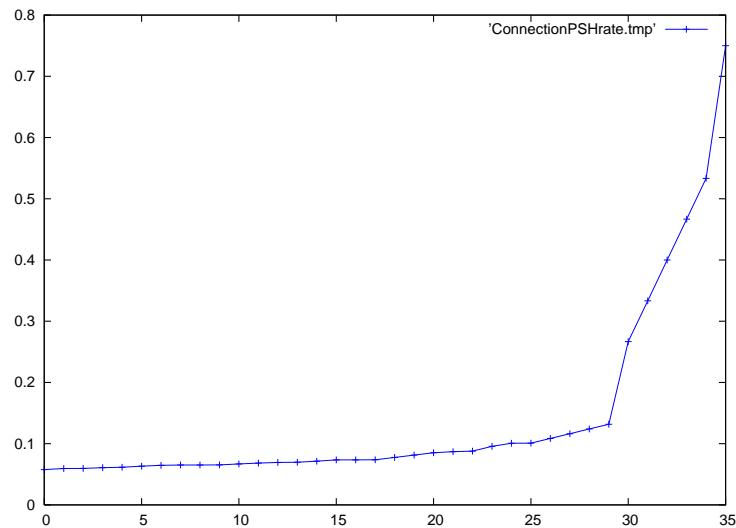


Figure H.45: Plot of number of connections versus the number of PSH flags, over the past m connections, for one hour of sample data.

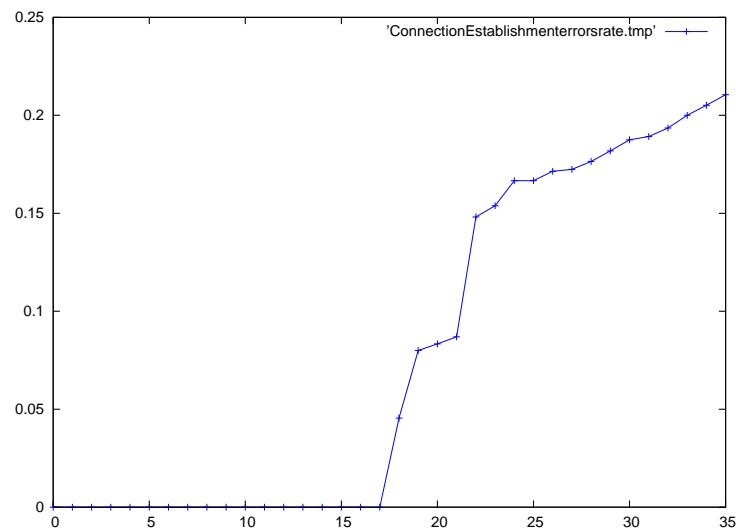


Figure H.46: Plot of number of establishment errors versus the number of connections, over the past m connections, for one hour of sample data.

is no upper bound to how high it might go. In the case of our sample data, plotted out in Figure H.46, all the establishment errors come from connections initiated to services that were not running on the sample network. Presumably, these connections were made in an attempt to gain some type of unauthorized access, although we have no way to prove the intent they were made with.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Other error versus connection, connection rate window

The category of “other errors” represents a number of TCP protocol violations that may occur while the connection is in the ESTABLISHED state, for example a packet with a SYN flag, or a packet without an ACK flag. Like establishment errors, there is technically no upper bound to this ratio. Such errors are rather blatant protocol violations, however, so we should only see them in the case of attacks or serious misconfigurations. Most of the time we expect this will be zero, as it is for our sample data as shown in Figure H.47.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Disconnection error versus connection, connection rate window

Similar to the other errors, disconnection errors are protocol violations that occur after at least one side has sent a FIN. There may be multiple disconnection errors per connection, and the ratio may exceed one. However, we only expect to see them in exceptional situations and most of the time, we expect the ratio to be at or near zero, as it is for our sample data as seen in Figure H.48.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

Average duration, connection rate window

Figure H.49 shows the average duration, in seconds, of the last m connections, plotted out in sorted order. There is no upper bound on these values, but realistically, one will not see any durations longer than the length of time the network has been observed for.

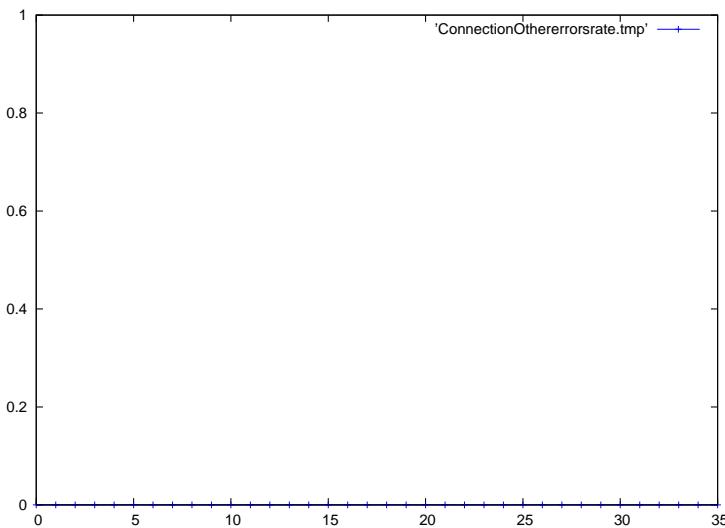


Figure H.47: Plot of number of other errors versus the number of connections, over the past m connections, for one hour of sample data.

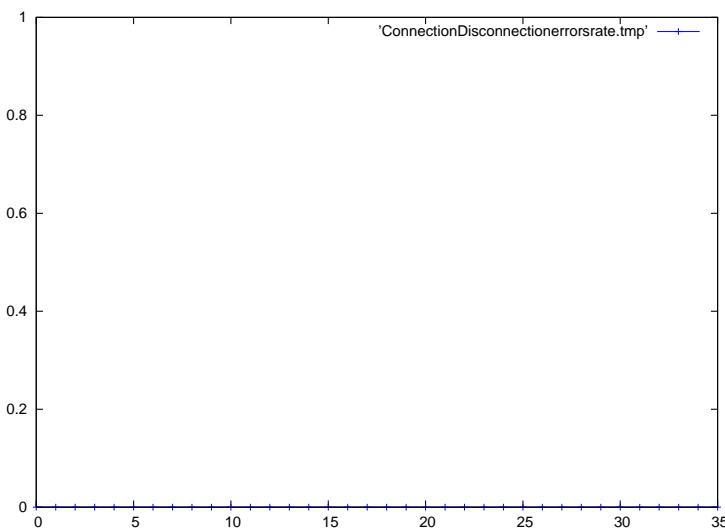


Figure H.48: Plot of disconnection errors versus the number of connections, over the past m connections, for one hour of sample data.

As with our other sorted ratio values, we think that scaling the graphs and comparing the points of the smaller graph to the possibly projected points of the larger graph will provide a good similarity metric.

H.1.6 Per connection close

We now take a look at the metrics of each individual connection, which we know at the time the connection closes.

Packets per connection

Figure H.50 shows the number of packets in each connection, plotted out with the number of packets along the x-axis, and the number of connections that had that many packets along the y-axis. We expected the initial bump for connections with only a few packets. What is particularly interesting is that there are multiple connections with the same number of packets in excess of 100, indicating that there is possibly a close relationship between the types of activity that take place on the network and the resulting number of packets in a connection.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

Packets sent per connection

Besides looking at the total number of packets per connection, we also want to look at how many of those were sent or received by the local network. Figure H.51 shows the number of packets sent from the local network per connection. Note that when we say “sent”, we just mean from the network that we are monitoring, without respect to which side initiated the connection. This means that the metric makes sense for connectionless protocols as well. Here we see a small spike towards the right that undoubtedly contributed to the spikes we saw to the right in Figure H.50.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

Packets received per connection

Figure H.52 shows the other side of the coin: the number of packets received by the local network per connection. Again, this is regardless of who initiated the connection. Discrepancies

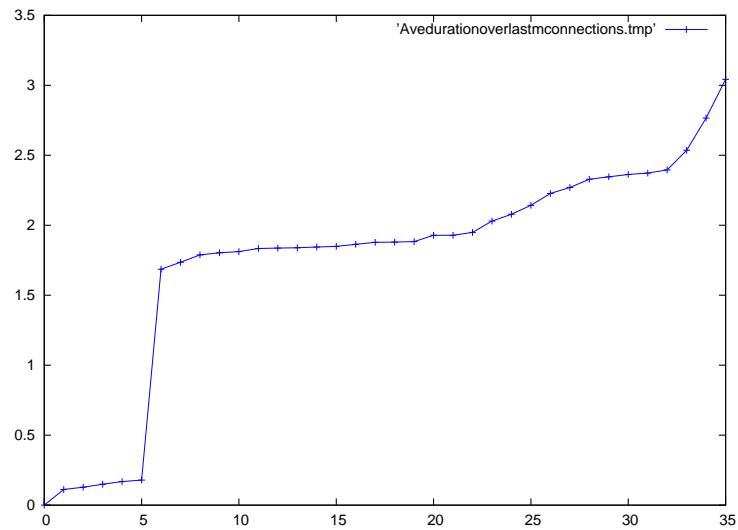


Figure H.49: Plot of average duration, over the past m connections, for one hour of sample data.

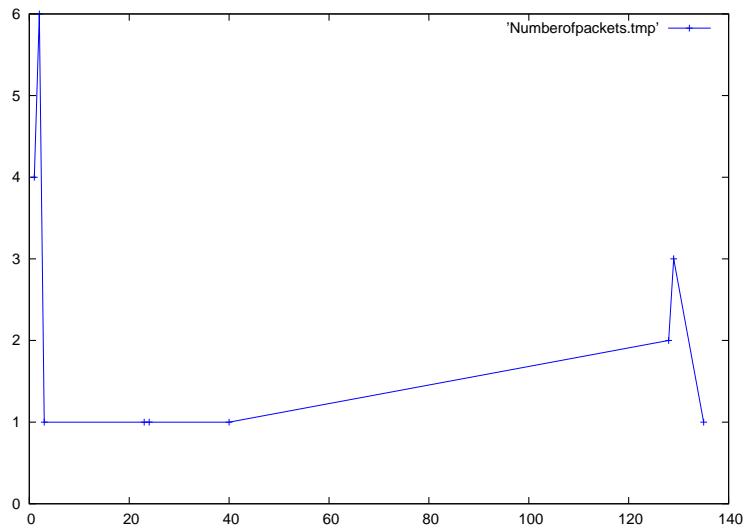


Figure H.50: Plot of packets per connection values for one hour of sample data.

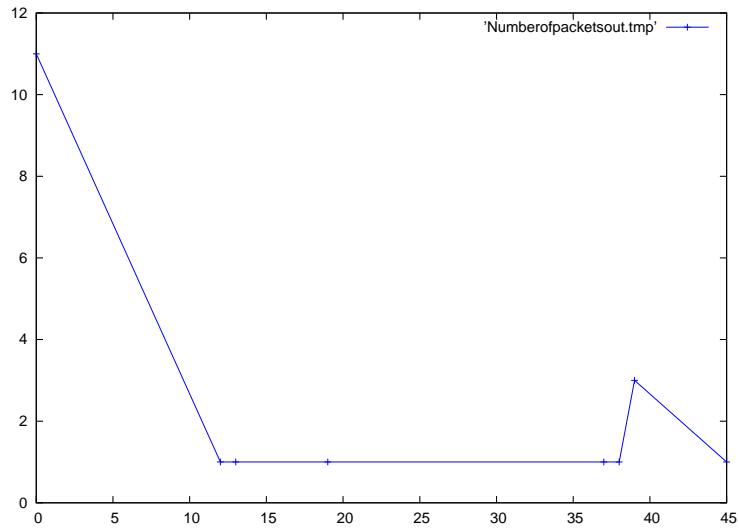


Figure H.51: Plot of packets sent per connection values for one hour of sample data.

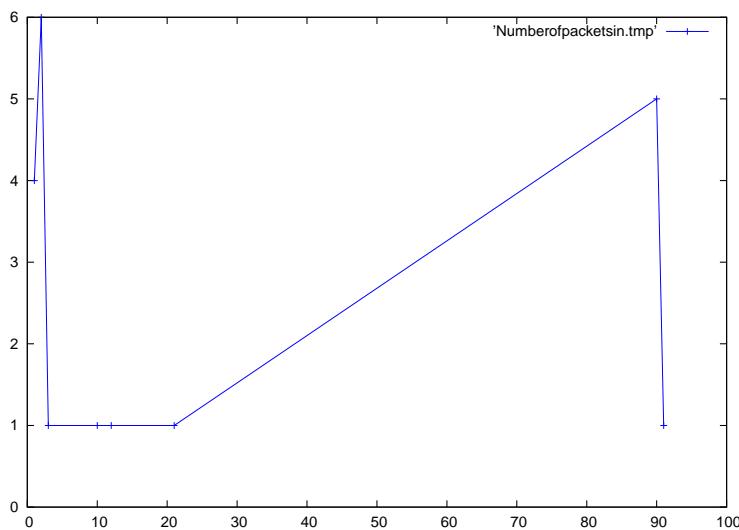


Figure H.52: Plot of packets received per connection values for one hour of sample data.

between this plot and Figure H.51 would be indicative of a client dominated or server dominated network, respectively. In this case, our sample network leans towards the client side. As we will discuss under the bytes and data bytes metrics, below, using packets sent or received is not a great metric of client or server activity because of ACK packets.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

Connection duration

Connection duration, as seen in Figure H.53, should provide some of the best insight into the usage of a given network, as the duration of connections is intuitively indicative of what they are used for: short connections (under 10 seconds) for activities such as web browsing, longer connections (in the 10's of seconds) for bulk transfers, such as software downloads, and long durations (100's of seconds or more) for interactive sessions. These metrics are the sort of thing that have been passed by oral tradition and are used in the qualitative comparison of networks. By actually comparing the connection duration curves of two traces, we will get to see if these qualitative metrics hold when doing a quantitative comparison.

Since we are not concerned with how the durations differ between adjacent connections, we have plotted them out in increasing order, as we did with packet and connection interarrival time. This leads to the intuition that the same comparison technique – scaling the x-axis of both graphs to the same scale and finding the mean similarity between each point on the smaller graph to a (possibly projected) point on the larger graph – will work well for comparing duration graphs.

Connection percent control packets

A control packet is one in which the data portion of the packet is zero. Put another way, a packet is either a data packet or a control packet, and the control packet rate is one minus the data packet rate. This can be seen in Figure H.54, which plots out the control packet rate of our sample data.

We plan to compare this like all of our other connection rates, however given that it is a perfect inverse of the data packet rate, the control packet rate similarity should be the same as the data packet rate similarity. We include it partially because some researchers used data packet rates and others used control packet rates, and partially as a sanity check to ensure that it really is the same.

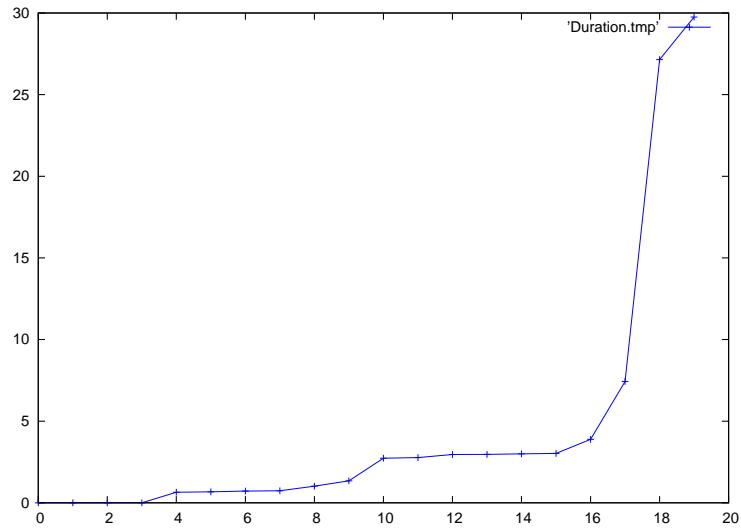


Figure H.53: Plot of connection duration values for one hour of sample data.

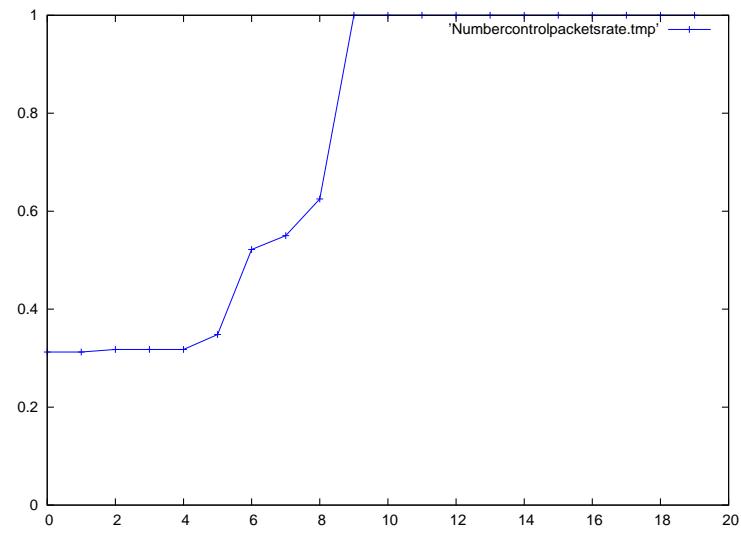


Figure H.54: Plot of connection percent control packets values for one hour of sample data.

Connection percent data packets

A data packet is one in which the data portion of the packet is greater than zero. The percentage of packets in a connection that are data packets is informative: if none of the packets were, this indicates that while there may have been the exchange of a large number of packets, no actual data was transferred, which rather defeats the purpose of making a connection; more likely it indicates a failure to connect. If about half the packets were data packets, it indicates a one-way exchange of data, such as in a file transfer where most of the packets in one direction are data packets, and most of the packets in the other direction are acknowledgments. Data packet rates well above half are more indicative of interactive connections, where a response packet with an acknowledgment also contains some data. Figure H.55 shows that about half the connections in our sample data involved no data transfer. Examining the data shows that these were suspicious connections that never made it to the establishment state. A few connections had data rates near 50% – indicative of file transfers. The remaining connections were in the 60% to 70% range – indicative of more interactive-type traffic.

Like our other connection rates, it appears that we can find the similarity between data packet rates by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

Bytes per connection

Above we looked at the number of packets per connection; now we turn to the finer grained metric of number of bytes per connection. As shown in Figure H.56, we did not get multiple connections with the exact same number of bytes, but we do see distinct clusters where the number of bytes were close. For this reason, we think that our comparison method on the weighted averages will continue to work well here.

Bytes sent per connection

As with the packets sent versus total packets per connection, we look at the bytes sent out from the observed network per connection to get a finer grained picture than what the total byte counts per connection gives us. As there, when we say “sent” we mean from the network we are monitoring, regardless of who initiated the connection. The result is shown in Figure H.57, which shows a significant spike at zero (this is likely all the connections that were ignored because they violated local policy), and an interesting spike of two connections to the right, which lends further credence that this will be a useful metric. We believe measures of this metric can be compared using the discrete comparison on the weighted averages.

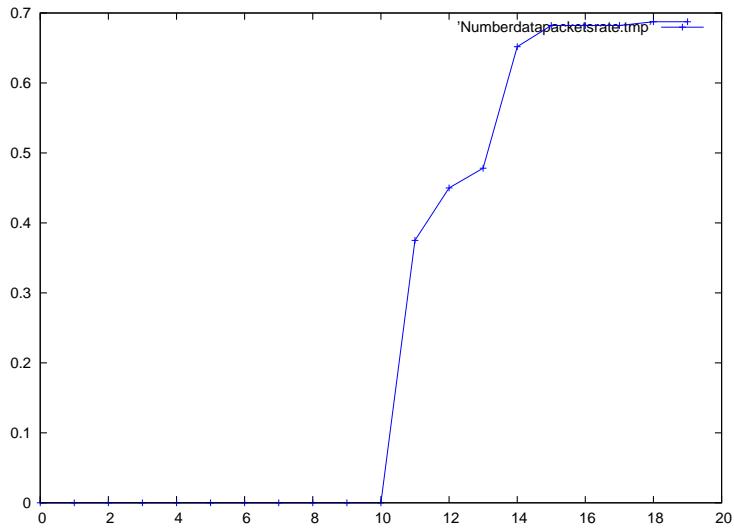


Figure H.55: Plot of connection percent data packets values for one hour of sample data.

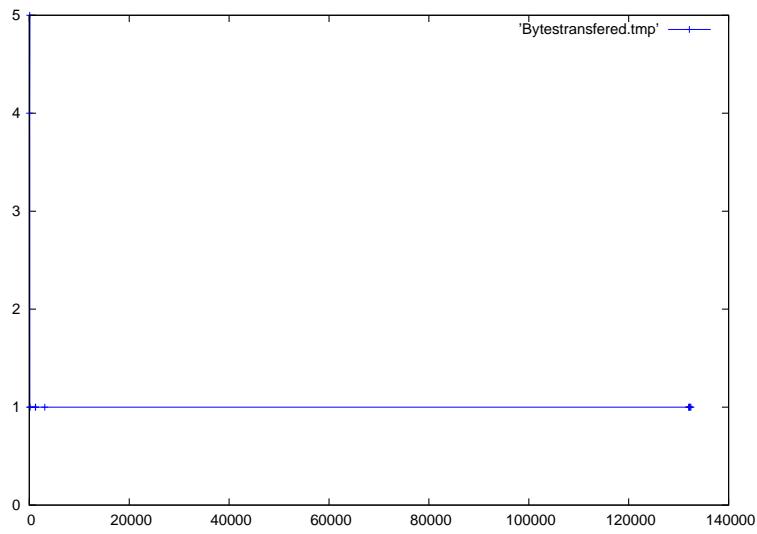


Figure H.56: Plot of bytes per connection values for one hour of sample data.

Bytes received per connection

The other side of the bytes sent per connection is the bytes received by the local network, as shown in Figure H.58, which looks remarkably similar to the total byte counts. At the same time, it looks fairly different from the bytes sent per connection, meaning that the total bytes per connection is likely dominated by the bytes received, which makes sense given the large values. This is probably a much better measure of the client versus server activity on the network than the packet counts (which include a lot of empty ACKs).

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

Data bytes per connection

Above we looked at the total bytes transferred per connection, which included the packet headers. Now we look at just the data portion of the connections. As Figure H.59 shows, this looks a lot like Figure H.56, except for the number of connections that have zero bytes transmitted. These, again, are likely the connections that never got established because they violated local policy. As such, we can already start to see that looking at the number of data bytes specifically is likely a very useful measure.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

Data bytes sent per connection

Figure H.60 really starts to give us a very concrete picture of the sample network, as it shows that none of the outgoing traffic transmitted more than 200 bytes per connection, with quite a few well below that.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

Data bytes received per connection

Figure H.61 shows the opposite side of the coin from Figure H.60. In fact, it looks almost identical to Figure H.59, as it should given the very small number of data bytes accounted for as outgoing. This metric, along with the data bytes sent per connection, firmly establishes that

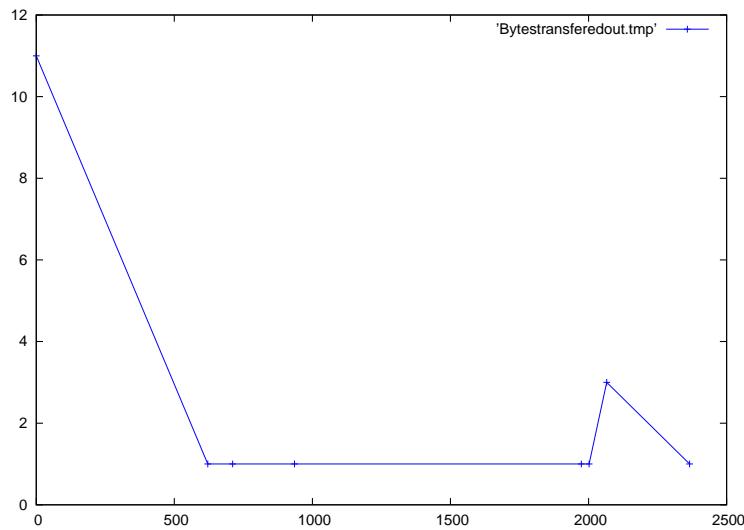


Figure H.57: Plot of bytes sent per connection values for one hour of sample data.

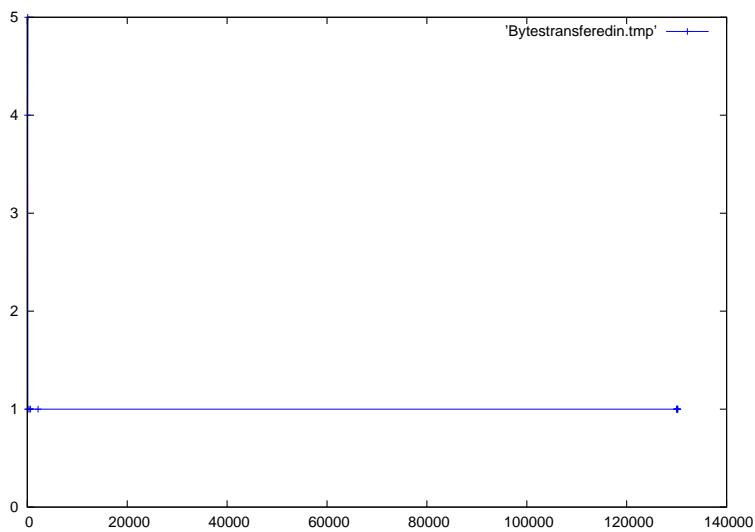


Figure H.58: Plot of bytes received per connection values for one hour of sample data.

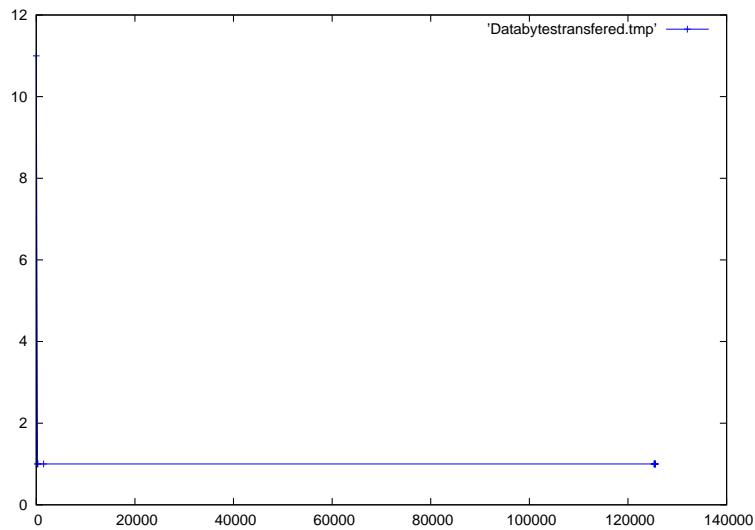


Figure H.59: Plot of data bytes per connection values for one hour of sample data.

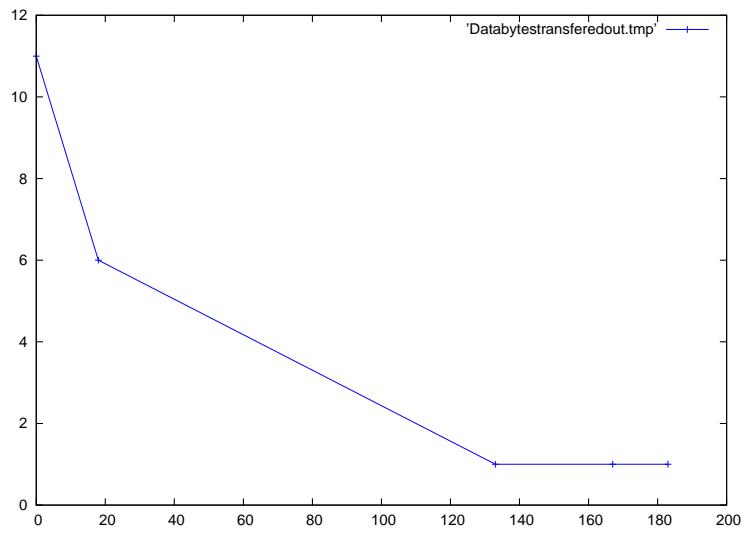


Figure H.60: Plot of data bytes sent per connection values for one hour of sample data.

the monitored network is primarily a receiver / consumer of data.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

Connection frag packet rate

As our one hour of sample data did not contain any fragmented packets, Figure H.62, which shows the cumulative connection fragmented packet rate values, is even more boring than the packets or connections per day of the week graphs. All we see is a line sitting on zero. Instead of relying on sample data, we will have to rely on our domain knowledge to know how to model this data.

One may ask that, if we did not even see any fragmentation in the sample data, why measure and compare on it at all? The reason is that on contemporary networks, fragmentation is typically not used at all, however it is useful for bypassing some network security devices such as primitive firewalls and many intrusion detection systems (Ptacek and Newsham 1998), so the presence of fragmentation is of interest to information security researchers. Further it should be of interest in comparing networks as the level of fragmentation on the network should indicate the level to which attackers are crafting their malicious traffic to bypass security devices. We expect to see more of such traffic on networks that are being explicitly targeted (which our network where the sample data comes from is apparently not).

Like connection duration and interarrival times, we are not concerned with the fragmentation rates on adjacent connections – rather we can plot the rates out in increasing order. This leads to the intuition that we can compare them like all the other metrics that we plot out in sorted order: by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

Connection wrong frag packet rate

Given the lack of any fragmentation, it should not be surprising that our sample data has no bad fragments, as shown in Figure H.63. As discussed in chapter 6, the wrong fragmentation packet rate is the percentage of packets in a connection that were fragmented incorrectly – that is, some fragments were overlapping, missing, or other similar fragmentation badness. Since we know that fragmentation is of interest from a security perspective, one can be assured that bad fragmentation is even more so.

Also unsurprisingly, we will use the same approach to model and compare the wrong fragmentation rate as we do the overall fragmentation rate: by finding the mean similarity

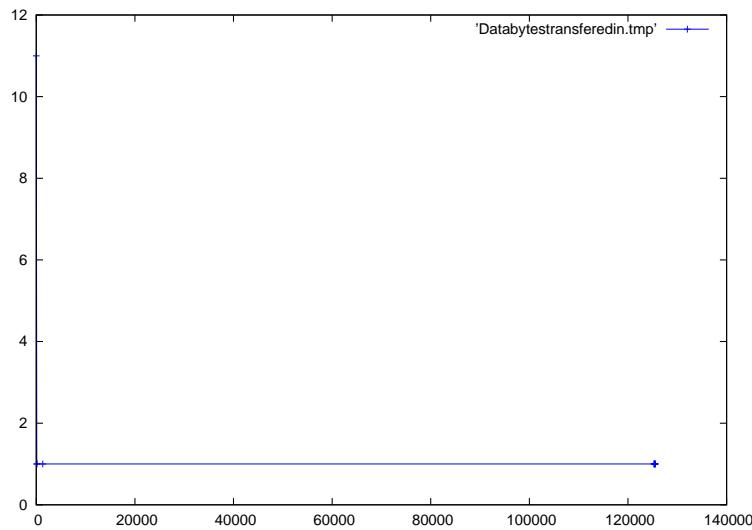


Figure H.61: Plot of data bytes received per connection values for one hour of sample data.

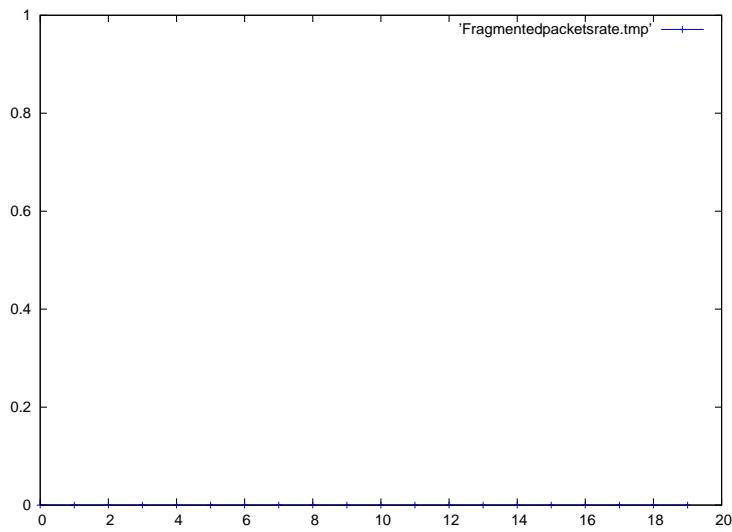


Figure H.62: Plot of connection frag packet rate values for one hour of sample data.

between every point on the smaller curve and a (possibly projected) point on the larger curve.

Connection source window size

Figure H.64 shows the plot of the max window sizes of the connection sources (x-axis) versus the number of connections that used that window size (y-axis). As one can see, these are all over the board, although most of the values are close to the maximum size this field allows (65536).

While we have used the same method as we used for the TTL and temporal characteristics to plot the window size values, we did not plot out all the zero values, which would otherwise overwhelm the plot. Likewise, if we used the same method to compare the values, the zeros would overwhelm the averages, especially if we extend out the size of the window for the weighted average, which we probably should since most of the non-zero values are separated by values on the order of thousands.

We also fear that part of our problem here is having such a small sample size, just as that complicated figuring out how to model TTL and packet interarrival times. Once again, we turn to our DSL data, as seen in Figure H.65. This is interesting because it shows the same general pattern, just more extremely pronounced. Put another way, there are about the same number of data points, and some of their values are much larger than in our one hour of sample data. Hence, we are left with the same conclusion that the method we used to model and compare TTL values will not work.

We began by modifying the approach we use for TTL: for each measurement (source window size, in this case), we compare its value against all the values corresponding to a measurement within 10 000 in the other trace. This value is then weighted linearly depending on how far the measurements are from one another (1 if they are equal, .5 if they are 5000 apart, 0 if they are 10 000 apart). The sum of these weighted similarities is divided by the sum of the weight to give the similarity for that measurement. The mean of these similarities is the overall similarity. In summary, if $meas1_{1..n}$ are the measurements in the first dataset, and $meas2_{i,1..m}$ are the measurements in the second dataset within 10,000 of $meas_i$, then we find the similarity by calculating:

$$\frac{\sum_{i=1}^n \sum_{j=1}^m \left(1 - \frac{|meas1_i - meas2_{i,j}|}{meas1_i + meas2_{i,j}}\right) \cdot \frac{|meas1_i - meas2_{i,j}|}{10000}}{\sum_{i=1}^n \sum_{j=1}^m \frac{|meas1_i - meas2_{i,j}|}{10000}}$$

$$n$$

This gives a reasonable similarity value when we compare the one hour of sample data to the DSL data. We noticed, however, that the value was not transitive: that is, if we reversed the order of the datasets, the similarity value was slightly different. This is most easily explained by considering that some measurements in the second dataset might not be within 10 000 spaces,

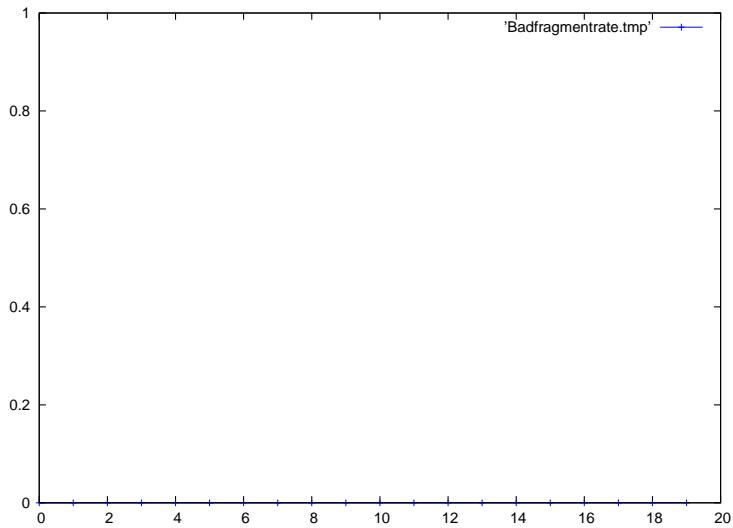


Figure H.63: Plot of connection wrong frag packet rate values for one hour of sample data.

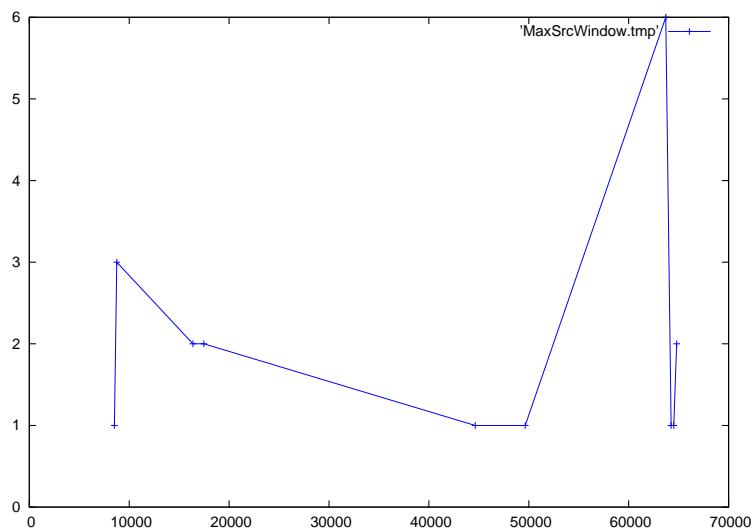


Figure H.64: Plot of connection source window size values for one hour of sample data.

and hence might not be compared against anything in the first dataset. More likely, even if a measurement in the second dataset was within 10 000 of a measurement in the first, if it is a significant distance away, the weighting factor will ensure that the similarity between the two does not carry much weight. While we say that we are trying to assign a quantitative value to the similarity between the two datasets, this similarity value should capture the differences between them as well, and if the second dataset has data points so askew from any points in the first dataset as to not significantly add to the similarity, we need to capture that as a dissimilarity.

We use a simple method to accomplish this: we run the formula as above, then we use the second dataset for *meas1* and the first dataset for *meas2*, and find the mean of the two similarities.

Connection destination window size

While the connection destination window size values, as seen in Figure H.66, do not look much like the values of the source window sizes, we can see that they are similar in that we are dealing with a small number of populated values separated by significant distance. As such, we can use the same comparison approach as we did for connection source window size.

Connection urgent packet rate

The TCP Urgent flag is meant to indicate that a given packet is important and should get preferential treatment in processing, such as being processed before other packets in the network stack. (Postal 1981) does not specify exactly what sort of preferential treatment should be given to such packets, and to the best of this author's knowledge, this feature is not used on contemporary networks. This serves to explain why Figure H.67, which shows the urgent packet rates of our connections, is so boring. In fact, the only time we expect to see URG set is when an attacker is attempting to exploit a flaw in the network stack, such as with a “Christmas tree packet” (Miller 2000).

Like the fragmentation rates, we will compare two sets of urgent packet rates by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

Connection resend packet rate

The resend rate is a good indicator of the reliability of connections to and from a given network. Since we are looking only at the connections in and out of a given network, what we are actually measuring here is the reliability rate for the connections that are being made into and from the network we are looking at. If a network has a close connection to the Internet

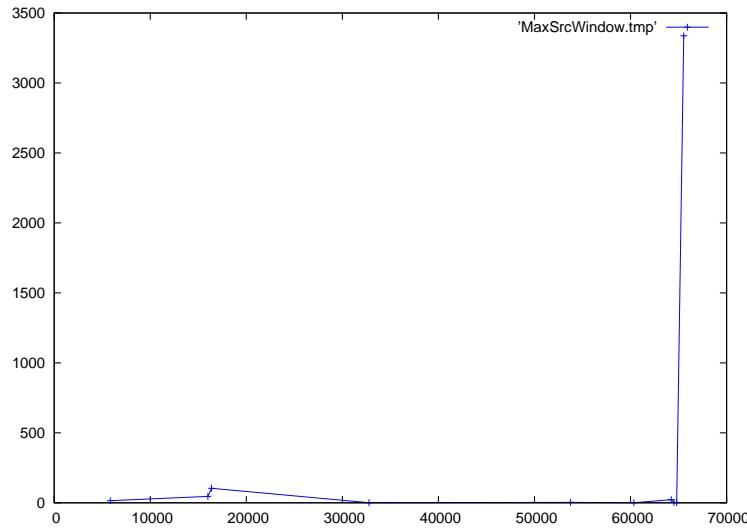


Figure H.65: Plot of connection source window size values for approx 2.5 days of DSL data.

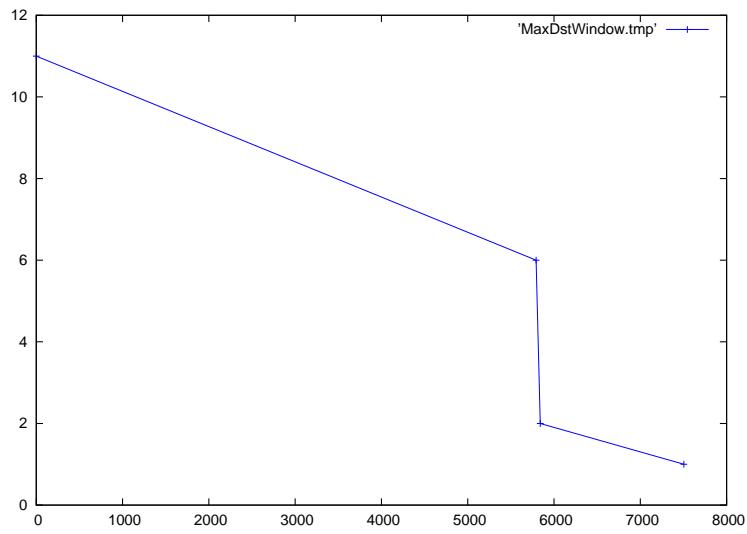


Figure H.66: Plot of connection destination window size values for one hour of sample data.

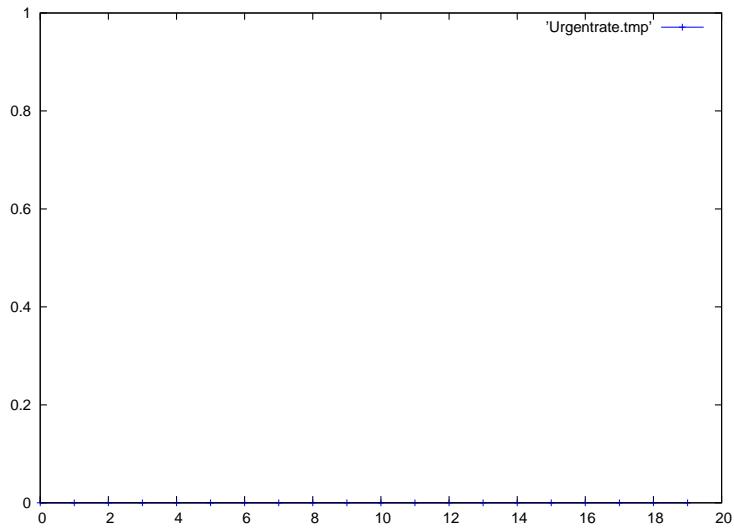


Figure H.67: Plot of connection urgent packet rate values for one hour of sample data.

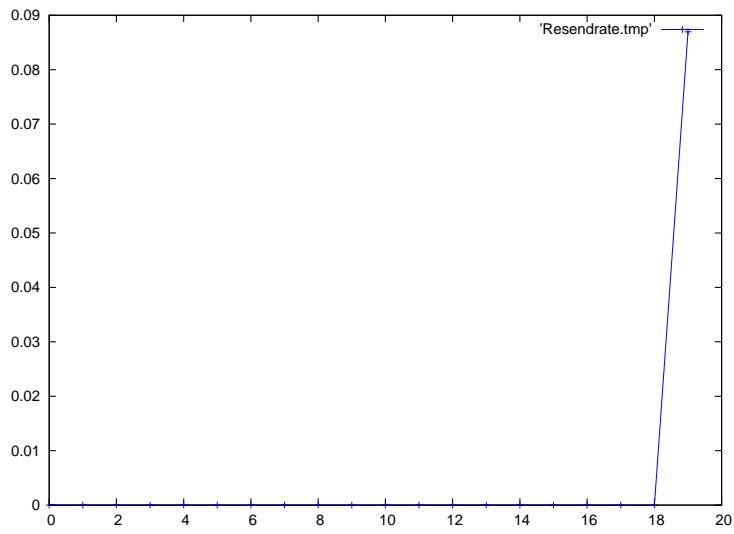


Figure H.68: Plot of connection resend packet rate values for one hour of sample data.

backbone and tends to only communicate with other highly reliable sites, we do not expect to see many resends. On the other hand, even if a network has a reliable Internet connection, if it gets a lot of traffic from clients on less reliable edge nodes, we should see an increase in the packet drop rate, which will manifest itself as higher resend rates. As Figure H.68 shows, most of the connections in our sample data had no resends, however one connection required that most of its packets were resent.

It appears that connection resend rates can be modeled and compared by the same method as all the other connection rates: by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

Connection wrong resend packet rate

Dropped packets and resends are a fact of life in packet switched networks. Wrong resends are a different matter. Wrong resends are when a packet is resend that should not be, like once its already been acknowledged. This may occur due to deficiencies in the network stack, or it may indicate malicious activity, such as a sequence injection attack. Granted, there may be occasions that an acknowledgment and the “wrong” resend cross paths on the network, but these are extremely rare, and indicate serious latency issues that the network stack should account for. As shown in Figure H.69, we do not see any wrong resends in our sample data.

Despite this lack of wrong resends in the sample data, our domain knowledge indicates that we can model and compare the wrong resend rates just like we do all the other connection rates: by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

Connection duplicate ACK packet rate

Since all packets in an established TCP connection ACK the last segment seen from the other host, it is typical to see the same segment acknowledged multiple times when one host is sending all the data. We should not, however, see a segment before the most recently acknowledged segment be reacknowledged. Nor should we see packets without any data be sent just to acknowledge a segment that has already been acknowledged. Seeing these kinds of duplicate acknowledgments indicates that something bad is happening on the network. Once again, we were not surprised to see that Figure H.70 shows no such activity in the sample data.

Despite this lack of duplicate ACKs in the sample data, our domain knowledge indicates that we can model and compare the duplicate ACK rates just like we do all the other connection rates: by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

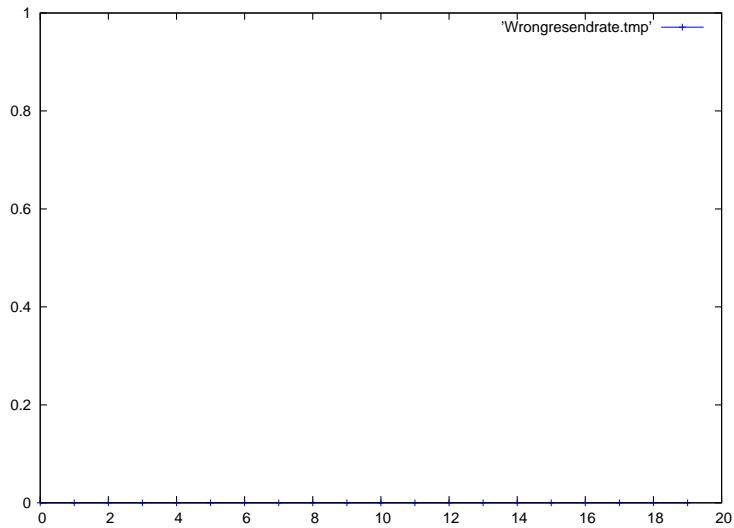


Figure H.69: Plot of connection wrong resend packet rate values for one hour of sample data.

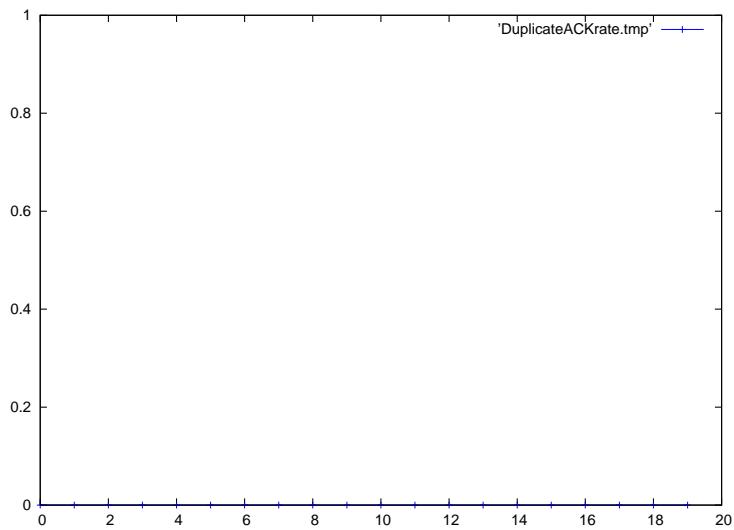


Figure H.70: Plot of connection duplicate ACK packet rate values for one hour of sample data.

Wrong ACK

While duplicate ACKs *might* be an indication that something is “hinky” on the network, a wrong ACK occurs when a segment is acknowledged that was never seen. This is definitely an indication that something is wrong, which in the best case is that the observer is not seeing some of the packets traversing the network. As Figure H.71 shows, unsurprisingly, there were no wrong ACKs in our sample data.

Despite this lack of wrong ACKs in the sample data, our domain knowledge indicates that we can model and compare the wrong ACK rates just like we do all the other connection rates: by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

Connection wrong data packet size rate

We consider packets that should not have data, but do, to have the wrong data packet size. Like many of our other metrics, this should only occur due to a network stack deficiency, or malicious activity. Once again, as Figure H.72 shows, there is none in the sample data.

Despite this lack of wrong data packet sizes in the sample data, our domain knowledge indicates that we can model and compare the wrong data packet size rates just like we do all the other connection rates: by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

Window exceeded rate

A TCP sender should know the window size of the receiver, and by keeping track of what segment has been acknowledged, and what the last segment sent was, keep from sending more data than the receiver can handle. In the event that the sender does send more than the receiver can handle, we call that a window exceeded error. As Figure H.73 shows, we did not have any of these in our sample data.

Despite this lack of window exceeded errors in the sample data, our domain knowledge indicates that we can model and compare the window exceeded error rates just like we do all the other connection rates: by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

Connection hole rate

A true connection hole indicates that something is seriously wrong on the network – this is when we see a segment acknowledged that was never received. If this actually occurs, it

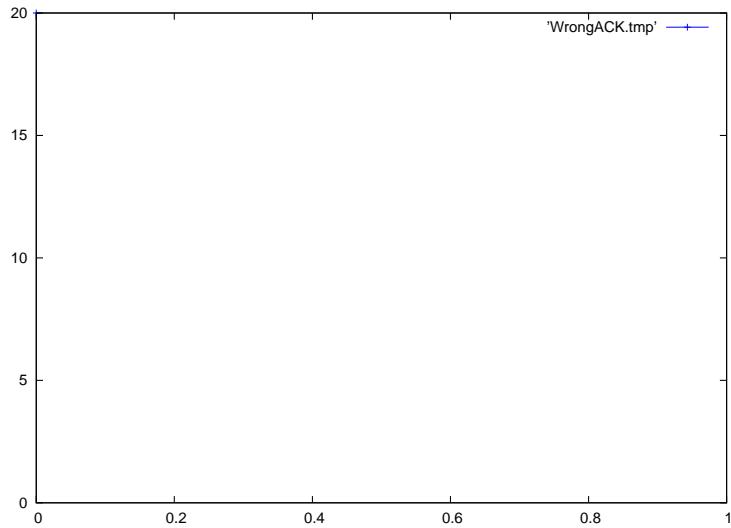


Figure H.71: Plot of wrong ACK errors per connection values for one hour of sample data.

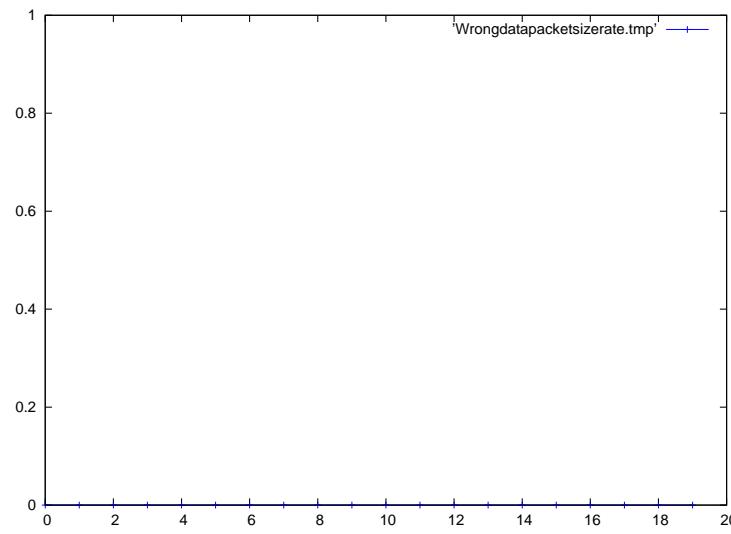


Figure H.72: Plot of connection wrong data packet size rate values for one hour of sample data.

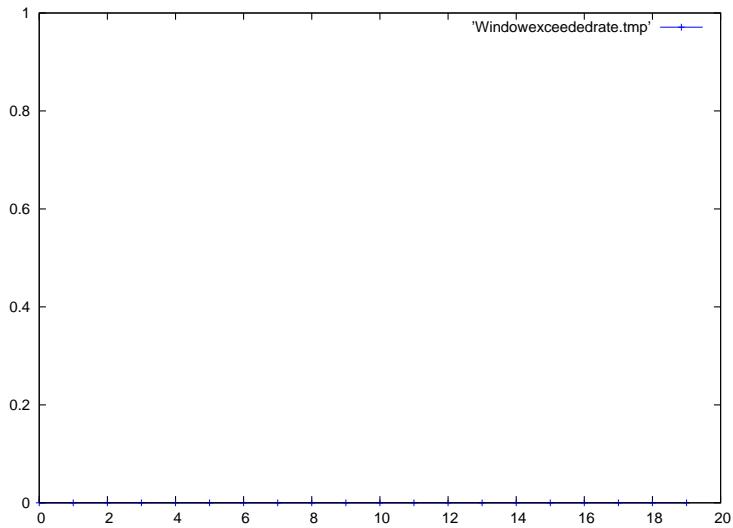


Figure H.73: Plot of window exceeded errors per connection values for one hour of sample data.

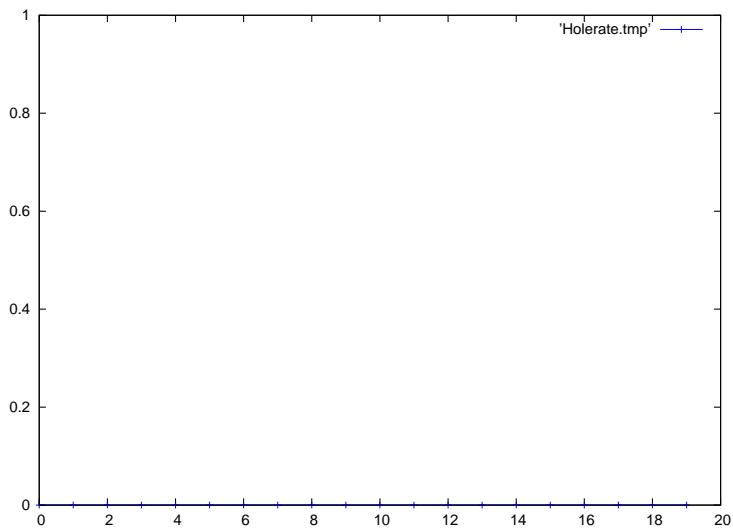


Figure H.74: Plot of connection hole rate values for one hour of sample data.

indicates a serious implementation problem with the acknowledger's network stack. More likely, it means that our trace collection is either dropping packets, or does not see all possible ingress and egress paths from the network. Once again, Figure H.74 shows that we do not have any such activity in our sample data.

Despite this lack of connection holes in the sample data, our domain knowledge indicates that we can model and compare the connection hole rates just like we do all the other connection rates: by finding the mean similarity between every point on the smaller curve and a (possibly projected) point on the larger curve.

Connection errors

Figure H.75 shows the number of connection errors for each connection. We were initially concerned that the data for this metric would not look like a continuous curve; fortunately, as one can see, it actually does looks continuous.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

Reset connection

Figure H.76 shows the number of resets seen per connection, as you can see, we did not have any in our sample data. Technically, there should only be one reset for any given connection, which would create a graph with only two values (zero and one). As discussed in chapter 6, we will group a series of resets with the same connection information together, meaning there is no limit to the number of resets we might see.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

Other errors

As discussed above, other errors are rare. We did not have any in our sample data as shown in Figure H.77. As with connection errors, we were initially uncertain if this would be modeled in a useful manner as a continuous characteristic, and given the apparent success with connection errors, we are now much more confident that this is the right approach for other errors.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing

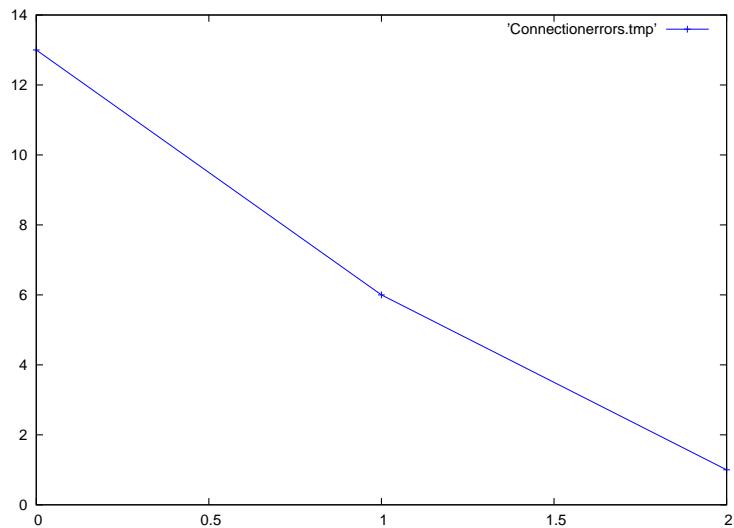


Figure H.75: Plot of connection errors values for one hour of sample data.

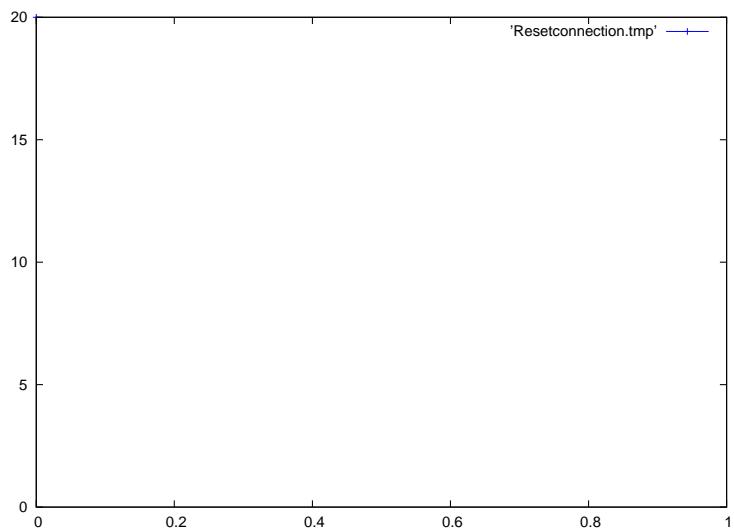


Figure H.76: Plot of reset connection values for one hour of sample data.

a discrete comparison on the weighted averages.

Disconnection errors

Figure H.78 shows the disconnection errors, or lack thereof, in the sample data. As with other errors, these are rare, but with the apparent success of modeling connection errors as continuous metrics, we believe this is the right approach for disconnection errors as well.

As with the other values that we look at by ordering along the x-axis and counting the number of connections along the y-axis, we anticipate that the best way to compare this is doing a discrete comparison on the weighted averages.

H.1.7 Per IP

Originally, we looked at the per IP metrics as discrete characteristics where the IP served as the independent variable and we looked at the number of connections, packets and bytes for each one. This approach works well when we have the actual IP addresses, as we can correlate activity to common hosts, such as popular Internet sites, or just track levels of activity to particular internal servers. The problem we encountered was when the observed network was anonymized – suddenly we could not key directly off the IP addresses. We worked around this by mapping the networks to each other to treat internal hosts that appeared to be similar as the same. This process is documented extensively in Appendix D. This worked decently well, although the algorithm employed was $O(n^2)$ where n was the number of hosts on the observed network. As such, this limited the size of the network that we could realistically compare. A second curve ball was thrown when we discovered that we would only realistically be able to obtain real network traces from most sources if they were completely anonymized: both the internal and external (or source and destination) addresses would be obfuscated. In this environment, trying to map IP addresses between traces seems far more tenuous and computationally expensive. As such, we rethought the approach. Instead of keying off of the IP addresses, we will simply build a distribution of the total number of connections, packets, or bytes per source and destination IP and look at it like any other sorted distribution.

It is worth noting that looking at these metrics *per IP* might be considered a second-order metric, which we were not going to look at during this stage. However, these are the only sensible ways to capture information about the source and destination IP addresses, so we need to consider these metrics in order to consider them for use in higher-order metrics (which we only build off of useful lower-order metrics).

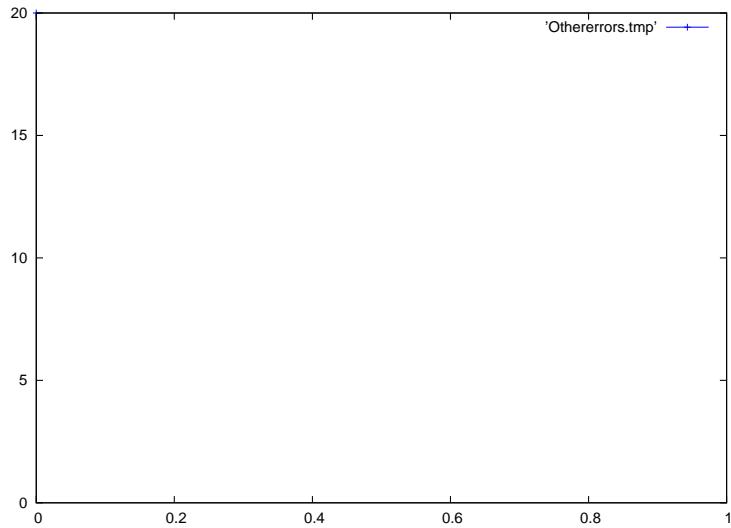


Figure H.77: Plot of other errors per connection values for one hour of sample data.

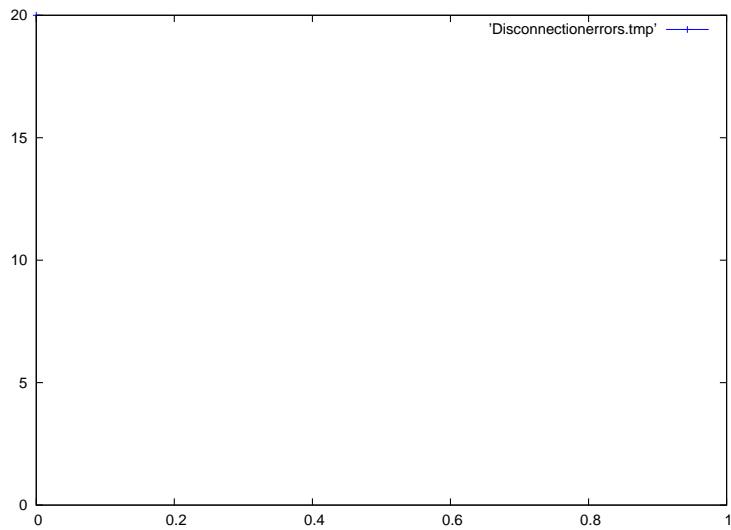


Figure H.78: Plot of disconnection errors values for one hour of sample data.

Number of connections per source IP

Figure H.79 shows the number of connections made per source IP, plotted out in sorted order. Most sources only made a single connection, with a few that made more, up to a max of 12 connections from a single source. The final point, with 12 connections, is the collocated host on the target network.

Since this data is conveniently represented as a curve of values sorted from lowest to highest – just like packet interarrival – we will use the same approach – comparing each value on the curve with less points to a possibly interpolated point on the larger curve.

When performing this comparison, we may want to consider scaling the y-axis, either on the basis of the maximum number of connections per source between the two traces, or on the basis of the total number of connections between the two traces. We are hesitant to do this, however, as, looking at the plot, it seems that most of the IP addresses have very few connections, and if we were to scale all the points with one connection to have a portion of a connection, that may have an adverse affect on the comparison. An alternative method of scaling might be to not scale by a constant factor, but do a logarithmic scaling to match where $y = 1$, and match the maximum number of connections per IP. For the time being, however, we will refrain from scaling and see if performance is acceptable. If it is, we will simply consider this for future work.

Number of connections per destination IP

Figure H.80 shows the number of connections made per destination IP, plotted out in sorted order. Note that there are many fewer destination IP addresses than source IP addresses, as many of the source IP addresses (particularly those with single connections) were probes that were not responded to. All of these had the destination IP of the collocated machine on the target network, accounting for the final point with a value of 24. The other three IP addresses were addresses that the target network desired communication with, most of them over multiple connections.

As with the other per IP metrics, we will compare the sorted curves scaled along the x-axis, but not along the y-axis.

Number of packets per source IP

Figure H.81 shows the number of packets sent per source IP, plotted out in sorted order. As before, the final point, with a value of 604, is the collocated machine on the target network. Interestingly, the other points on the curve work progressively up to that point, as opposed to it representing a significant jump in the values.

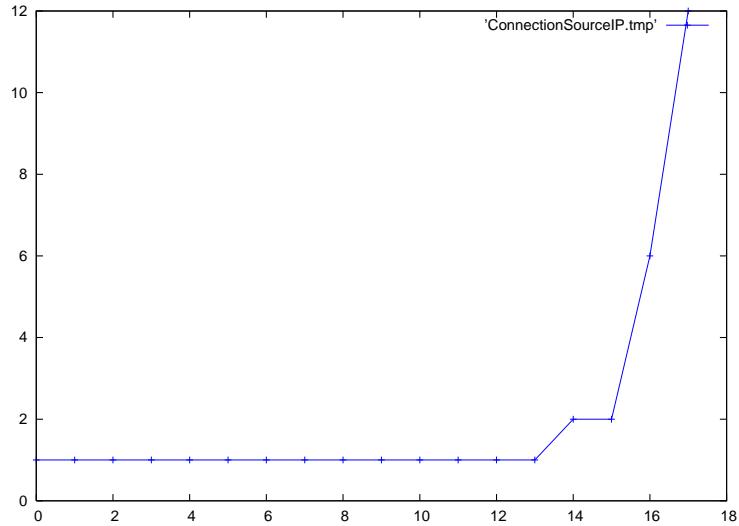


Figure H.79: Plot of number of connections per source IP for one hour of sample data.

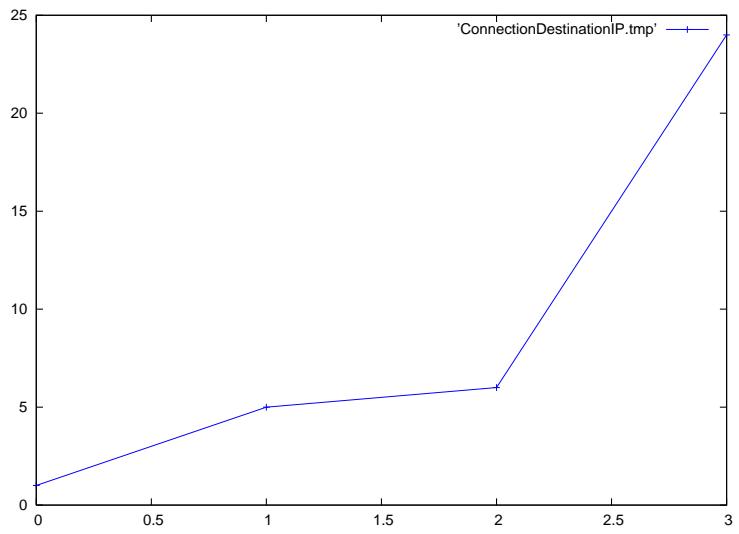


Figure H.80: Plot of number of connections per destination IP for one hour of sample data.

As with the other per IP metrics, we will compare the sorted curves scaled along the x-axis, but not along the y-axis.

Number of packets per destination IP

Figure H.82 shows the number of packets received per destination IP, plotted out in sorted order. As before, the final point, with a value of 922, is the collocated machine on the target network. Note that we have more addresses represented here than we did for the number of connections per destination IP, as we include the destinations of every packet observed here, whereas we only include the connections that we have complete information on in that metric.

As with the other per IP metrics, we will compare the sorted curves scaled along the x-axis, but not along the y-axis.

Number of bytes transferred per source IP

Figure H.83 shows the number of bytes sent per source IP, plotted out in sorted order. In an interesting switch, the collocated machine on the target network is the third from the last point. While this is useful information as it demonstrates that the target network acts much more in a client capacity than a server capacity. Unfortunately, these semantics are not captured in the data as represented. It may be worth considering separating our “Per IP” metrics into separate metrics for the internal and external IP addresses (with respect to the target network); this may be considered for future work.

As with the other per IP metrics, we will compare the sorted curves scaled along the x-axis, but not along the y-axis.

Number of bytes transferred per destination IP

Figure H.84 shows the number of bytes received per destination IP, plotted out in sorted order. The final point, with a value of 846380, is once again the collocated machine on the target network. Unlike the previous plots, the final point in this plot shows a significant jump above the previous value, again demonstrating that the target network is more of a consumer than producer of information.

As with the other per IP metrics, we will compare the sorted curves scaled along the x-axis, but not along the y-axis.

H.2 Normalized Similarity

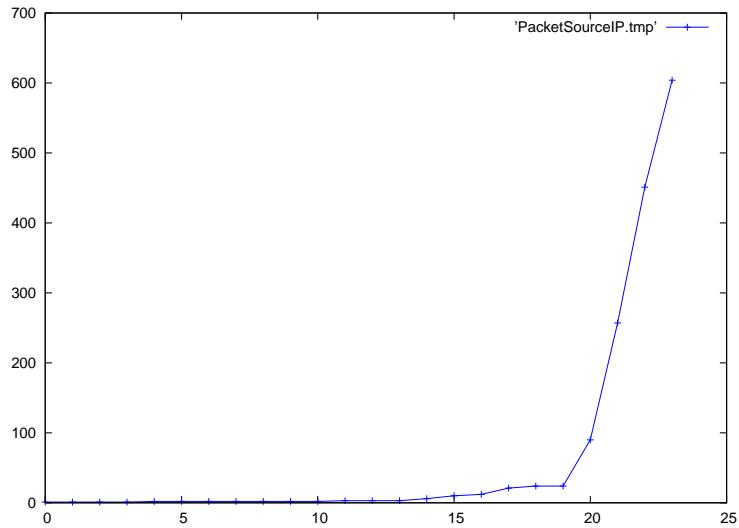


Figure H.81: Plot of number of packets per source IP for one hour of sample data.

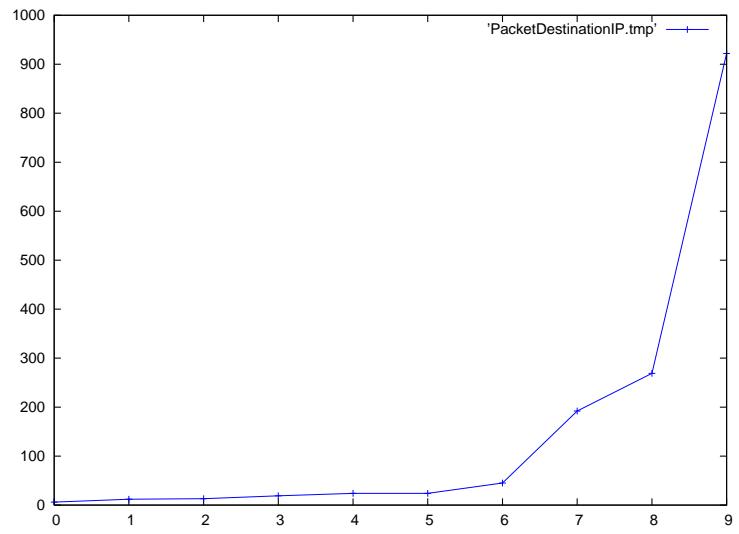


Figure H.82: Plot of number of packets per destination IP for one hour of sample data.

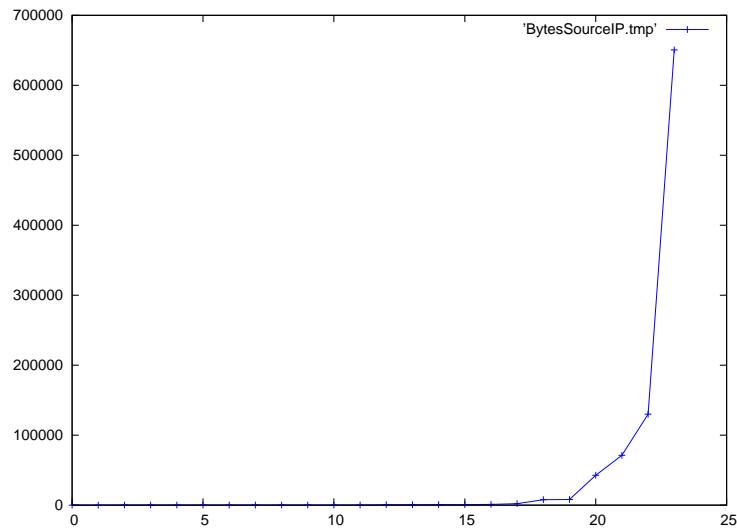


Figure H.83: Plot of number of bytes transferred per source IP for one hour of sample data.

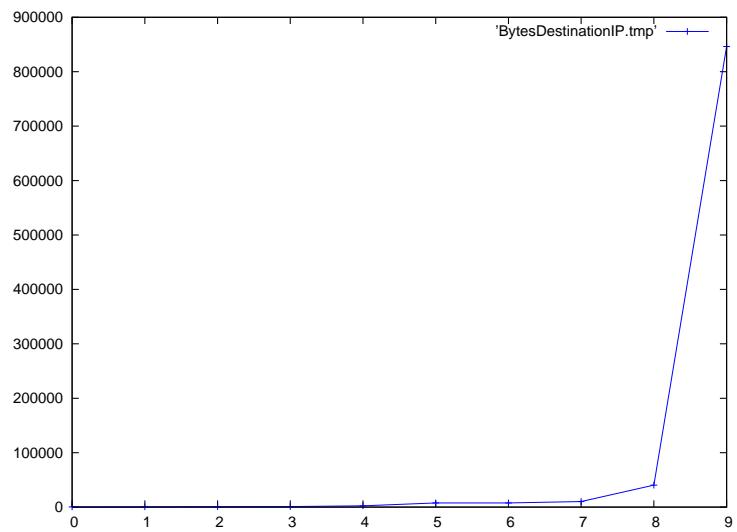


Figure H.84: Plot of number of bytes transferred per destination IP for one hour of sample data.

Table H.1: The normalized similarities of all characteristics for basecases 1 through 5.

Characteristic	Basecase 1	Basecase 2	Basecase 3	Basecase 4	Basecase 5
Packets in	1.0	1.0	0.9989	0.9978	0.1376
Packets out	1.0	1.0	0.9992	0.9967	0.1131
Connections in	1.0	1.0	0.9796	0.96	0.15
Connections out	1.0	1.0	1.0	1.0	0.1519
Bytes in	1.0	1.0	0.9994	0.9995	0.0151
Bytes out	1.0	1.0	0.9994	0.9506	0.0011
SYN-ONLY rate	1.0	1.0	0.9780	1.0	0.8903
SYN-ACK rate	1.0	1.0	1.0	1.0	1.0
Idle connection rate	1.0	1.0	1.0	1.0	0.0
Half-open connection rate	1.0	1.0	0.9934	1.0	0.8889
Packet Service	1.0	1.0	0.8883	0.7190	0.0
Bytes Service	1.0	1.0	0.8911	0.7150	0.0
Connection Service	1.0	1.0	0.8940	0.8281	0.0
Packet Source port	1.0	1.0	0.8941	0.7513	0.0
Bytes Source port	1.0	1.0	0.8937	0.7499	0.0
Connection Source port discrete	1.0	1.0	0.8454	0.7459	0.0
Connection Source port continuous	1.0	1.0	0.9999	0.9999	0.9963
Packet TTL discrete	1.0	1.0	0.7959	0.6401	0.5637
Packet TTL continuous	1.0	1.0	0.9858	0.9741	0.8720
InterPacket delta	1.0	1.0	0.9637	0.9146	0.2581
Packet sec	1.0	1.0	0.6655	0.5705	0.1169
Packet min	1.0	1.0	0.9836	0.8200	0.1230
Packet GmHour	1.0	0.8611	0.8611	0.8611	0.8181
Packet LocHour	1.0	0.8611	0.8611	0.8611	0.8181
Continued on next page					

Table H.1 – continued from previous page

Characteristic	Basecase 1	Basecase 2	Basecase 3	Basecase 4	Basecase 5
Packet weekday	1.0	1.0	0.9998	0.9985	0.5010
Bytes sec	1.0	1.0	0.6414	0.4860	0.0059
Bytes min	1.0	1.0	0.9947	0.8511	0.0069
Bytes GmHour	1.0	0.8611	0.8611	0.8611	0.7932
Bytes LocHour	1.0	0.8611	0.8611	0.8611	0.7932
Bytes weekday	1.0	1.0	0.9997	0.9974	0.4328
Size	1.0	1.0	0.9947	0.9688	0.6035
Packets in last w secs	1.0	1.0	0.9880	0.9212	0.0823
Priv packets time rate	1.0	1.0	0.9915	0.9349	0.4620
Unpriv packets time rate	1.0	1.0	0.9962	0.9701	0.9303
Connections time rate	1.0	1.0	0.9866	0.9271	0.7120
Priv connections connection time rate	1.0	1.0	0.9923	0.9776	0.4748
Unpriv connections connection time rate	1.0	1.0	0.9969	0.9469	0.6395
Priv packets priv connection time rate	1.0	1.0	0.9756	0.9451	0.4426
Unpriv packets unpriv connection time rate	1.0	1.0	0.9935	0.9008	0.6487
SYNs connection time rate	1.0	1.0	0.9797	0.8450	0.6432
RSTs connection time rate	1.0	1.0	0.9989	0.9980	0.9602
FINs connection time rate	1.0	1.0	0.9988	0.9947	0.5162
PSH connection time rate	1.0	1.0	0.9888	0.9407	0.5881
Establishment errors connection time rate	1.0	1.0	0.9902	0.9421	0.1923

Continued on next page

Table H.1 – continued from previous page

Characteristic	Basecase 1	Basecase 2	Basecase 3	Basecase 4	Basecase 5
Other errors connection time rate	1.0	1.0	1.0	1.0	1.0
Disconnection errors connection time rate	1.0	1.0	1.0	1.0	1.0
Ave duration over last w secs	1.0	1.0	0.9681	0.7145	0.4459
Priv packets packet rate	1.0	1.0	0.9958	0.9801	0.7630
Unpriv packets packet rate	1.0	1.0	0.9993	0.9960	0.9481
InterConnection delta	1.0	1.0	0.8722	0.8389	0.1429
Connection sec	1.0	1.0	0.7216	0.5922	0.1559
Connection min	1.0	1.0	0.9429	0.8656	0.1602
Connection GmHour	1.0	0.8611	0.8611	0.8611	0.8230
Connection LocHour	1.0	0.8611	0.8611	0.8611	0.8230
Connection weekday	1.0	1.0	0.9922	0.9846	0.5146
Connection packet rate	1.0	1.0	0.9889	0.9898	0.5603
Connection Priv connections rate	1.0	1.0	0.9913	0.9846	0.9441
Connection Unpriv connections rate	1.0	1.0	0.9923	0.9819	0.9381
Connection Priv packet rate	1.0	1.0	0.9461	0.9607	0.3034
Connection Unpriv packet rate	1.0	1.0	0.9911	0.9852	0.8688
Connection SYNs rate	1.0	1.0	0.9332	0.9150	0.8626
Connection RSTs rate	1.0	1.0	0.9647	0.8002	0.5875
Connection FINs rate	1.0	1.0	0.9191	0.9527	0.8388
Connection PSH rate	1.0	1.0	0.9741	0.9703	0.2994
Continued on next page					

Table H.1 – continued from previous page

Characteristic	Basecase 1	Basecase 2	Basecase 3	Basecase 4	Basecase 5
Connection Establishment errors rate	1.0	1.0	0.9353	0.9294	0.1889
Connection Other errors rate	1.0	1.0	1.0	1.0	1.0
Connection Disconnection errors rate	1.0	1.0	1.0	1.0	1.0
Ave duration over last m connections	1.0	1.0	0.9089	0.8876	0.6615
Number of packets	1.0	1.0	0.9834	0.9216	0.6869
Number of packets in	1.0	1.0	0.9610	0.8697	0.6060
Number of packets out	1.0	1.0	0.9529	0.8435	0.4073
Duration	1.0	1.0	0.8849	0.8866	0.7858
Number control packets rate	1.0	1.0	0.9835	0.9936	0.5624
Number data packets rate	1.0	1.0	0.9747	0.9915	0.6534
Number bytes transferred	1.0	1.0	0.9999	0.9998	0.9993
Number bytes transferred in	1.0	1.0	0.9999	0.9998	0.9977
Number bytes transferred out	1.0	1.0	0.9971	0.9929	0.9998
Number data bytes transferred	1.0	1.0	0.9999	0.9998	0.9993
Number data bytes transferred in	1.0	1.0	0.9999	0.9998	0.9977
Number data bytes transferred out	1.0	1.0	0.9738	0.9339	0.9998
Fragmented packets rate	1.0	1.0	1.0	1.0	0.75
Continued on next page					

Table H.1 – continued from previous page

Characteristic	Basecase 1	Basecase 2	Basecase 3	Basecase 4	Basecase 5
Bad fragment rate	1.0	1.0	1.0	1.0	0.75
Max Src Window	1.0	1.0	0.9999	0.9997	0.9991
Max Dst Window	1.0	1.0	1.0000	0.9995	0.9982
Urgent rate	1.0	1.0	1.0	1.0	1.0
Resend rate	1.0	1.0	0.9474	0.95	0.6587
Wrong resend rate	1.0	1.0	1.0	1.0	0.7
Duplicate ACK rate	1.0	1.0	1.0	1.0	1.0
Wrong ACK	1.0	1.0	1.0	1.0	1.0
Wrong data packet size rate	1.0	1.0	1.0	1.0	0.0
Window exceeded rate	1.0	1.0	1.0	1.0	0.7
Hole rate	1.0	1.0	1.0	1.0	1.0
Number connection er- rors	1.0	1.0	0.9559	0.9657	0.9624
Reset connection	1.0	1.0	0.9607	0.9730	0.8827
Other errors	1.0	1.0	0.9744	1.0	0.1869
Disconnection errors	1.0	1.0	0.9607	0.9730	0.8827
Packet Destination IP	1.0	1.0	0.9959	0.9710	0.7188
Bytes Destination IP	1.0	1.0	0.9932	0.9801	0.0180
Connection Destination IP	1.0	1.0	0.9949	0.99	0.8458
Packet Source IP	1.0	1.0	0.9962	0.9745	0.8655
Bytes Source IP	1.0	1.0	0.9508	0.9301	0.1087
Connection Source IP	1.0	1.0	0.9815	0.9778	0.9725

H.3 Scaled Similarity

Table H.2: The weighting factors for all characteristics found with linear regression, using the goal similarities of $\{1.0, 1.0, 0.9, 0.8, 0.1\}$.

Characteristic	wf
Packets in	0.0104
Packets out	0.0101
Connections in	0.0112
Connections out	0.0105
Bytes in	0.009
Bytes out	0.0098
SYN-ONLY rate	0.057
SYN-ACK rate	0.0
Idle connection rate	0.0089
Half-open connection rate	0.074
Packet Service	0.0128
Bytes Service	0.0129
Connection Service	0.0109
Packet Source port	0.0123
Bytes Source port	0.0123
Connection Source port discrete	0.0118
Connection Source port continuous	2.5183
Packet TTL discrete	0.0295
Packet TTL continuous	0.0893
InterPacket delta	0.0142
Packet sec	0.0134
Packet min	0.0142
Packet GmHour	0.0
Packet LocHour	0.0
Packet weekday	0.0179
Bytes sec	0.0126
Bytes min	0.0117
Bytes GmHour	0.0
Continued on next page	

Table H.2 – continued from previous page

Characteristic	wf
Bytes LocHour	0.0
Bytes weekday	0.0158
Size	0.0258
Packets in last w secs	0.0113
Priv packets time rate	0.0204
Unpriv packets time rate	0.2187
Connections time rate	0.0451
Priv connections connection time rate	0.0182
Unpriv connections connection time rate	0.032
Priv packets priv connection time rate	0.0185
Unpriv packets unpriv connection time rate	0.0389
SYNs connection time rate	0.0428
RSTs connection time rate	0.2397
FINs connection time rate	0.0188
PSH connection time rate	0.0274
Establishment errors connection time rate	0.0125
Other errors connection time rate	0.0
Disconnection errors connection time rate	0.0
Ave duration over last w secs	0.0287
Priv packets packet rate	0.0434
Unpriv packets packet rate	0.1961
InterConnection delta	0.0125
Connection sec	0.015
Connection min	0.0133
Connection GmHour	0.0
Connection LocHour	0.0
Connection weekday	0.0192
Connection packet rate	0.0206
Connection Priv connections rate	0.2178
Connection Unpriv connections rate	0.2073
Connection Priv packet rate	0.013
Continued on next page	

Table H.2 – continued from previous page

Characteristic	wf
Connection Unpriv packet rate	0.0784
Connection SYNs rate	0.0828
Connection RSTs rate	0.0377
Connection FINs rate	0.0396
Connection PSH rate	0.0133
Connection Establishment errors rate	0.0119
Connection Other errors rate	0.0
Connection Disconnection errors rate	0.0
Ave duration over last m connections	0.0339
Number of packets	0.0411
Number of packets in	0.0347
Number of packets out	0.0217
Duration	0.0438
Number control packets rate	0.02
Number data packets rate	0.0246
Number bytes transferred	18.0213
Number bytes transferred in	4.2184
Number bytes transferred out	1.2766
Number data bytes transferred	18.2326
Number data bytes transferred in	4.2317
Number data bytes transferred out	0.1325
Fragmented packets rate	0.0356
Bad fragment rate	0.0356
Max Src Window	15.2947
Max Dst Window	7.4708
Urgent rate	0.0
Resend rate	0.0288
Wrong resend rate	0.0297
Duplicate ACK rate	0.0
Wrong ACK	0.0
Wrong data packet size rate	0.0089
Continued on next page	

Table H.2 – continued from previous page

Characteristic	wf
Window exceeded rate	0.0297
Hole rate	0.0
Number connection errors	0.07
Reset connection	0.0722
Other errors	0.0105
Disconnection errors	0.0722
Packet Destination IP	0.038
Bytes Destination IP	0.0094
Connection Destination IP	0.0634
Packet Source IP	0.0899
Bytes Source IP	0.011
Connection Source IP	0.3433
y-intercept	-73.8785

Appendix I

Notes on real data

Chapter 8 provides an overview of the available network traces, along with the datasets that were selected for our testing on real data. That selection process was driven by a much more in-depth examination of the data than was presented in the overview. This appendix provides the notes we compiled on the various data sources, divided by data source.

I.1 Internet Traffic Archive

Nothing useful in the Internet Traffic Archive (ITA), as all of its traces have been converted from pcap to ASCII format.

I.2 UNC/FORTH

Can not use UNC/FORTH packet header traces as they have been converted into an ASCII format that is specific for wireless research.

I.3 MAWI

MAWI Working Group Traffic Archive (WIDE project) - Japan/US peering point

- Daily samples (not continuous) from 23 Feb 1999 to present
- <http://tracer.csl.sony.co.jp/mawi/>
- Trace from 2007-11-07

- 15 minutes
- 33,146,641 packets
- ten nets above 1,000,000 packets:

203.112.243	3190826
89. 62. 81	2830817
164. 92.104	2430097
177. 9.213	2103948
3.156.101	1742035
203.122. 92	1151863
203.112.246	1128049
133.108.232	1058421
202. 56.138	1058127
163.119.141	1046463

- Trace from 2007-11-08
 - 15 minutes
 - 29,198,977 packets
 - six nets above 1,000,000 packets:

110. 65.199	2729148
155.154. 84	2252811
72.141. 76	1662651
196.209.189	1288628
164.163. 87	1096399
197. 19.196	1085423

I.4 Crawdad

Crawdad - archive of wireless network data, available at <http://crawdad.cs.dartmouth.edu/data.php>, include acknowledgment: “We gratefully acknowledge the use of wireless data from the CRAWDAD archive at Dartmouth College.”

I.4.1 Dataset: `dartmouth/campus`

- fall01: 2001-09-20 – 2002-01-10
 - AcadBldg16
 - * 2001-09-25 – 2002-01-08 (each trace covers 24 hours)
 - * 4,208,693 IP packets in 24 hours
 - * Local Net: 190.84.172
 - SocBldg1

- * 2001-09-25 – 2002-01-10
- * 391,927 IP packets in 24 hours
- * Local Net: 190.84.200
- spring02: 2002-03-21 – 2002-06-09
- fall03: 2003-11-01 – 2004-02-27
 - LibBldg1-Sniffer1 local nets

190.84.117 1512347
190.84.105 2138470
190.84.116 4207422
 - LibBldg1-Sniffer2 local nets

190.84.116 2551738
190.84.117 3200452
 - LibBldg1-Sniffer3 local nets

190.84.116 870226
190.84.117 1728697

I.4.2 Dataset: microsoft/osdi2006

- 5–8MB chunks covering a few minutes each
- 10 collection points (S1 to S10), nine wireless plus WAN interface
- Collection points do not necessarily map to access points
- Some collection points monitored two channels – presented as “A” and “B”
- S1/A* files
 - 2006-11-06 – 2006-11-07 (9 hours from each of two days)
 - 6,034,878 IP packets from 1st day
 - No clear local net – none over 500000, 19 over 100000
 - Anonymization appears completely random
- S10 is traffic btw wireless and WAN
- 2GB data

I.4.3 Dataset: pdx/vwave

- seconds field on each packet corrupted (all other data seems reasonable)
- snaplen may be too short for entire (TCP/UDP/ICMP)/IP header
- psu-cs
 - 2006-06-12 - 1 hour
 - 68,059 IP packets
 - Local nets:

251.146.212	49191
251.146.227	71683
- library
 - 2006-07-10
- cafeteria
 - 2006-07-10
- pioneer-sq
 - 2006-06-26
- urban-grind
 - 2006-07-20 - 2 hours
 - 393,640 IP packets
 - Local net: 10.52.15: 403190
- powells
 - 2006-07-24

I.4.4 Dataset: ucsb/ietf2005/wireless

- 2005-03-09 – 2005-03-10
- 3 channels, day and plenary
- first dump has 1 hour of data, plenary has 1 hour more

- day 1
 - 512,899 IP packets
 - Local nets:

130.129.135	184277
130.129.132	176166
243. 21.194	107211
- day 1 plenary
 - 390,383 IP packets
 - Local nets:

130.129.128	2637
130.129.129	13912
130.129.130	42347
130.129.131	19803
130.129.132	56904
130.129.133	35363
130.129.134	37086
130.129.135	168,838

I.4.5 Dataset: ucsd/sigcomm2001

- 2001-08-29 – 2001-08-31
- 16,370,446 IP packets
 - but only 15,302 SYNs
 - all but 179 of those are to 64.159.221
 - over half of those are from 64.0.0.0/8
- no dominate source network – looks more like a CoLo or server network than a conference network.
- do not recognize the ports – are they anonymized as well?
- 10636195 non-TCP packets
 - 8429856 of those not between 64.144.219.109 and 109.233.111.207
- Nets with most traffic:

109.233.111	9497656
68.144.219	8745717
64.159.221	7086498

- 64.0.0.0/8 seems to have a lot of source networks, but most of them go to 64.159.221.0/24.
- is 64.159.221.0/24 some sort of management net? Is not the source of any outgoing connections.
- Only 179 SYNs not involving 64.159.221.0/24 (out of 15302), and all of those involve 68.144.0.0/16.
- Overall, there do not seem to be any source networks for outgoing connections. This does not look like a client network as we would expect to find at a conference. It looks more like a server network, particularly for a server network in a corporate environment where the servers are hit predominately by other networks in the company. In any case, given the inconsistencies between what the dataset claims to be and what it appears to be, it is not suitable for our purposes.

I.4.6 Dataset: uw/sigcomm2004

- 2004-08-30 – 2004-09-03
 - 2004-08-31 – 2004-09-03
 - 8,458,614 IP packets
 - Local nets:

98. 93.251	5748331
252.167.255	2849505
- wireless samples
 - 2004-08-30 – 2004-08-31
 - 3,661,439 packets
 - Local net: 103.191.7: 2375590

I.5 UMass trace repository

Available at <http://traces.cs.umass.edu/index.php/Network/Network>

I.5.1 Dataset: uprm-wireless

- 25 + 22 traces
- ranging for 0 to 1132189 IP packets
- from a few seconds up to 13 minutes on 2005-01-24
- comparing two different WiFi systems in a variety of settings
- wireless_traces
 - 13 tests
 - 7 with data from a single node
 - 5 with data from two nodes
 - Last test looks like just the data from the second node (based on name)
 - each dump is just two nodes communicating
 - all traffic apparently between client and tcp/5001
 - Each node in each test has six trace files of a few minutes each
 - Last two tests have one additional dump per node (different channel?)
 - Packet counts:

Min	1,613
Mean	699,869.594594595
Max	1,093,621
Total	77,685,525

I.5.2 Dataset: web ident

- 2003-10 – 2004-03
 - October
 - 14342 trace files of a few seconds each (more than one flow each)
 - all traffic apparently between client and tcp/8888 (not tcp/113)
 - all traffic to and from localhost (127.0.0.1) (anonymization technique?)
 - Packet counts:

Min	5
Mean	841.768581787756
Max	3,815
Total	12,072,645

- 2006-02 – 2006-04
 - 51 directories of 2000 traces each
 - 2006-02-10 – 2006-03-15
 - all traffic apparently between client and tcp/22 (not tcp/113)
 - all traffic between 10.10.10.10 and 10.10.10.11 (anonymization technique?)
 - actually have separate source and dest
 - no SYNs?
 - Packet counts:

Min	0
Mean	289.314068627451
Max	28802
Total	29510035

I.6 UNC Mobile

Available at <http://www.cs.unc.edu/Research/mobile/datatraces.htm>; covers 2005-04-13 – 2005-04-20; No good: pcap dumps converted into ASCII format with much data lost.

I.7 CAIDA (DatCat)

Available at <http://www.datcat.org>

I.7.1 Dataset: AMES Internet Exchange (OC-48)

- 2002-08-14 16:00 – 19:00
 - First five minutes:
 - * 57,167,586 packets
 - * 214,249 class C subnets
 - * Six class-Cs above 1,000,000 packets:

0.	3.117	3837772
238.109.212		1998248
3.	3.241	1917534
3.236.223		1437433
162.131.189		1332597
241.	46.218	1118966
 - Second five minutes:

- * 55,861,008 packets
- * 214,412 class C subnets
- * Six class-Cs above 1,000,000 packets:

0.	3.117	3498540
238.	109.212	2115163
3.	3.241	1849187
	3.236.223	1438002
	162.131.189	1268216
241.	46.218	1076134

- 2003-01-15
- 2003-04-24 07:00 – 2003-04-24 08:00
 - 202511729 packets

I.7.2 Dataset: passive-2007/ampath-oc12

- 2007-01-08-23:00 – 2007-01-11-01:00
- first hour
 - 6,872,338 packets
 - 9041 class-C subnets
 - Eight class-Cs above 500,000 packets:

104.	16.126	1538804
22.	241. 98	1538356
56.	97. 78	1053559
100.	146. 0	1000100
106.	95.114	812893
97.	59.202	767014
106.	95.125	591030
98.	38.132	537078
 - last hour
 - 7,110,694 packets
 - 12,287 class-C subnets
 - Five class-Cs above 500,000 packets:

106.95.124	744506	
111.	2. 99	727031
106.95.114	561246	
125.64.101	515711	
97.59.202	514521	

- last 30 seconds
 - 57,444 packets
 - 1082 class C subnets
 - top five class-C subnets above are the top five again:

106.95.124	7390
111. 2. 99	5524
106.95.114	6035
125.64.101	4176
97.59.202	4935

- Packet headers appear to either be corrupt or – more likely (given that tcpdump can process it) – using a non-standard pcap format. My tools think that the first packet has a length of -9 (should not the length field be unsigned?) and ipsumdump crashes quickly when attempting to process the files as well. Tried on both big-endian and little-endian machines, so the host byte order is not the issue.

I.7.3 Dataset: SOTM27

- 2003-03-01 – 2003-03-06
 - miscDumps/sotm27.tcpdump
 - 54,536 packets
 - 172.16.134.0 local net

I.7.4 Dataset: Numerous DNS root server traces

- 2006-01-10 – 2006-01-11
- restricted (contract)

I.7.5 Dataset: KAIST-KOREN 1 Gigabit Ethernet Trace

- 2006-01-10 – 2006-01-11
- restricted (they will run the code)

I.7.6 Dataset: LBNL

- Traces are grouped to cover multiple (but not necessarily all) ports on a router in groups between 2004-10-04 – 2005-01-08. Collects taken on:
 - 2004-10-04
 - 2004-12-15
 - 2004-12-16
 - 2005-01-06
 - 2005-01-07
- Each trace appears to be from a different interface on a large switch or router connecting LBNL/UCB to the rest of the world. As such, it captures both traffic between local networks and the WAN as well as traffic between local nets. The router appears to cover 128.3.0.0/16 and 131.243.0.0/16. There is also a lot of traffic with 128.55.0.0/16, however all of that is to either of the other two networks, so this is likely either a Berkeley network handled by a different router, or the network of an organization that many at Berkeley interact with. Since each port on the router sees both local and external traffic, it is hard to determine what subnets are behind each port, and which are just communicating with nets behind that port. As such, the analysis below notes what networks are “externally facing”, meaning that we saw non-multicast traffic (which may in fact be local) between them and networks outside 128.3.0.0/16 and 131.243.0.0/16, pretty much ensuring that those nets are behind the given ports. The remaining traffic we call “internally facing” and consists of those nets that only had multicast traffic and/or communicated with other 128.3.0.0/16 and 131.243.0.0/16 nets. These internally facing nets are ranked by traffic volume, on the basis that we should see more traffic from networks that are actually behind a given port, until the networks, as presented, account for all the traffic.
- Each “trace” below is actually a pair of traces: the main trace and a trace of the detected scanning traffic for the same port over the same time period. We have recombined these two traces into a single trace using `ipsumdump`(Kohler 2004), for the purposes of this analysis.
- First trace
 - 20:03–20:13
 - port 1
 - 84574 + 1154 packets

- Only 33 multicast packets
- and 597 UDP packets (less than 1%)
- externally facing local nets:

128.3. 47	42446
128.3. 44	38653
128.3. 45	2511
128.3.100	1099
128.3. 46	436

- best internal facing local nets:

128. 3. 96	770
128. 3.148	708
128. 3. 99	511
128. 3.209	312
131.243.130	141
131.243.133	112
131.243. 94	86
131.243. 63	70

- Second trace

- 20:05–20:15
- port 2
- 5,703,128 packets
- Does not appear to monitor the same nets as the previous trace
- Appears to cover completely different networks: one different 128.3.X.0/24 and a bunch of 131.243.X.0/24s.
- A lot more UDP traffic (59%)
- 7358 Multicast packets (less than 1%)

- externally facing local nets:

128. 3. 23	5691253	131.243. 39	48	131.243. 68	4
131.243.199	1038	131.243. 33	40	131.243.130	4
131.243.108	303	131.243.208	24	131.243.112	4
131.243. 18	111	131.243. 87	19	131.243. 10	2
131.243. 4	99	131.243.251	18	131.243. 50	2
131.243.230	93	131.243.232	15	131.243. 65	2
131.243.253	86	131.243.234	4	131.243.118	2
131.243.139	60	131.243. 99	4	131.243. 51	2
131.243. 13	52	131.243.119	4	131.243. 58	1

- best internal facing local nets:

128. 3. 70	3403002	128. 3.204	41	131.243. 62	12
128. 3. 2	1811	131.243.125	41	131.243.144	12
128. 3. 96	1536	131.243.149	40	131.243.145	12
128. 3. 71	1145	131.243.127	38	131.243. 90	11
131.243. 63	826	131.243. 91	37	131.243.102	11
131.243. 61	374	131.243. 95	37	131.243. 15	10
128. 3.183	355	128. 3. 63	34	131.243.101	8
131.243.103	286	128. 3.148	34	131.243. 86	8
128. 3.255	259	131.243.106	34	131.243.147	8
131.243. 26	188	131.243.235	32	128. 3.210	8
131.243. 94	187	128. 3.211	32	128. 3. 37	8
131.243. 88	182	131.243. 14	32	128. 3.190	7
131.243.141	134	128. 3. 18	30	128. 3.147	6
128. 3. 99	128	131.243. 12	30	131.243.126	6
131.243. 85	83	131.243.104	24	128. 3. 6	5
128. 3.191	83	131.243. 89	21	131.243.203	4
131.243. 93	78	128. 3. 7	21	128. 3. 41	4
131.243.124	77	131.243.170	20	131.243.155	4
131.243.140	76	128. 3.189	20	131.243. 41	4
128. 3. 5	70	131.243.105	18	128. 3. 38	4
131.243.143	69	131.243. 92	17	131.243.181	4
131.243.142	68	131.243.146	16	128. 3. 39	4
128. 3.242	65	128. 3. 19	16	128. 3.174	2
128. 3.209	65	128. 3.188	15	131.243. 42	2
128. 3.208	59	131.243.100	15	131.243. 40	1
128. 3.194	49	131.243. 84	14	131.243.201	1
131.243.219	46	131.243.107	13		
128. 3.112	44	128. 3. 4	13		

- Third trace
 - 20:13–20:23
 - port 3
 - 2,261,261 packets
 - Appears to watch different specific 131.243.X.0/24s (131.243.40.0/22?) than either of the above.
 - 1–2% UDP traffic
 - Less than 1% multicast traffic
 - externally facing local nets:

131.243.140	1780053
131.243.142	185311
131.243.141	181806
131.243.143	60645
 - best internal facing local nets:

131.243. 95	574229	131.243. 41	999	128. 3. 7	86
128. 3.255	229036	131.243. 92	991	131.243.144	85
131.243. 84	175392	131.243. 61	969	131.243. 89	73
131.243.102	146222	128. 3.147	896	128. 3. 63	63
128. 3. 70	137659	131.243.104	809	131.243. 90	60
128. 3.209	137154	131.243.203	654	131.243.235	47
131.243.160	122446	128. 3.194	620	131.243.106	46
128. 3. 19	105457	128. 3.148	578	128. 3.210	44
128. 3.208	44012	131.243. 10	512	131.243. 14	43
131.243.170	25192	131.243.201	484	128. 3. 39	40
128. 3.189	23140	128. 3. 99	408	128. 3. 37	38
128. 3.211	22941	131.243. 60	379	131.243.105	37
128. 3. 18	21184	131.243. 94	289	131.243.208	36
131.243. 63	18811	131.243.232	262	131.243.145	34
128. 3.188	17029	131.243.101	235	128. 3. 6	34
131.243.202	13450	128. 3. 43	222	131.243. 15	32
131.243. 86	7416	131.243.124	199	128. 3. 5	22
131.243. 62	5259	128. 3. 48	160	131.243. 12	22
131.243.127	4062	131.243.146	147	131.243. 13	20
131.243. 87	3310	128. 3.191	146	128. 3.193	12
128. 3. 71	3091	131.243. 91	121	131.243. 42	12
131.243.103	2414	128. 3.204	119	131.243. 40	11
128. 3.242	2155	131.243.107	117	128. 3. 23	10
131.243. 26	1853	131.243. 88	109	128. 3. 4	9
131.243. 85	1730	131.243. 93	100	128. 3. 40	7
131.243.100	1547	128. 3.112	98	128. 3. 41	6
131.243.219	1456	131.243.147	96	131.243.126	4
128. 3.183	1059	131.243.155	91	128. 3.205	3

- Fourth trace

- 20:16–20:26
- port 4
- 5051 packets
- Again, different 128.3.X.0/24 subnet than any of the above.
- 1934 UDP packets (38%)
- 132 Multicast packets (2–3%)
- externally facing local nets:

128.3.26	3835
----------	------

- best internal facing local nets:

131.243.208	328
128. 3. 99	324
131.243. 63	76
131.243. 27	3

- Port 3 2004-12-15 11:09–12:09

- externally facing local nets:

131.243.160	1844847	131.243.111	76	131.243.	4	2
131.243.142	1318745	131.243.	80	66	131.243.	27
131.243.141	508339	131.243.	53	55	131.243.	186
131.243.143	270666	131.243.	132	52	131.243.	1
131.243.140	215987	131.243.	73	50	131.243.	66
128. 3. 2	6580	131.243.	185	42	131.243.	175
131.243.146	2597	131.243.	190	40	131.243.	159
131.243.135	1805	131.243.	67	38	131.243.	233
128. 3.148	1380	131.243.	226	38	131.243.	222
128. 3. 99	1295	131.243.	72	32	131.243.	230
131.243. 85	1291	131.243.	215	32	131.243.	193
128. 3. 97	1041	131.243.	181	29	131.243.	68
131.243. 24	672	131.243.	228	26	131.243.	19
131.243. 90	621	131.243.	251	26	131.243.	69
131.243.246	189	131.243.	45	26	131.243.	206
131.243.232	163	131.243.	210	25	131.243.	43
131.243.201	152	131.243.	52	6	131.243.	223
131.243. 94	128	131.243.	184	4	131.243.	82
131.243.199	108	131.243.	174	2	131.243.	212
131.243. 12	100	131.243.	48	2	131.243.	29
131.243.238	76	131.243.	156	2	131.243.	173

- best internal facing local nets:

128. 3.188	99949	131.243.103	576	131.243.124	138
128. 3.211	18697	131.243. 93	540	131.243.106	125
131.243. 26	3772	128. 3.193	529	131.243. 14	124
131.243. 63	3408	131.243. 41	474	131.243.125	102
128. 3.147	2904	131.243. 61	433	128. 3. 40	96
131.243.170	2064	131.243. 89	421	128. 3.183	94
128. 3.255	1792	131.243. 92	411	131.243.234	89
128. 3. 48	1602	128. 3. 19	409	131.243.155	88
131.243.100	1536	128. 3. 43	364	131.243. 13	84
128. 3.204	1280	128. 3.210	348	131.243.107	76
131.243. 95	1235	128. 3. 23	343	131.243.147	74
131.243. 84	1211	128. 3.242	342	131.243.104	68
131.243. 10	1122	128. 3. 39	333	131.243.105	66
131.243.102	1083	128. 3.190	327	128. 3. 38	52
128. 3.209	955	131.243. 60	312	128. 3. 37	50
128. 3. 41	930	131.243. 15	304	128. 3. 5	46
131.243.219	880	131.243. 91	300	131.243. 40	42
128. 3.189	835	131.243. 87	215	128. 3. 6	38
128. 3.208	825	128. 3.194	182	128. 3. 70	30
131.243.145	748	131.243.101	172	131.243.126	30
131.243. 62	658	131.243.144	166	128. 3. 4	30
131.243.203	655	131.243. 86	161	128. 3. 36	28
128. 3. 78	605	128. 3.112	160	131.243. 42	16
131.243.208	596	128. 3. 7	155	128. 3. 71	4
131.243. 88	588	131.243.127	144	128. 3.174	4
128. 3.191	587	128. 3. 63	141	128. 3.205	1

- Port 3 2004-12-15 19:42–20:42

– externally facing local nets:

131.243.140	6936110	128. 3.161	308630
131.243.160	1401278	128. 3. 2	12407
131.243.141	1043864	128. 3.148	5845
131.243.142	761237	128. 3. 99	4783
131.243.143	519468	131.243. 24	654

– best internal facing local nets:

128. 3.255	2591826	131.243.232	3926	131.243. 91	727
131.243. 95	1266494	131.243.102	3733	128. 3. 43	641
128. 3. 7	357350	131.243. 93	3686	131.243. 41	580
128. 3. 71	324280	128. 3.194	3655	131.243. 12	456
131.243. 61	311150	128. 3.183	3643	131.243. 15	448
128. 3. 19	228457	131.243.100	3564	131.243.155	346
128. 3. 18	227753	128. 3.204	3066	131.243.124	327
128. 3.147	154734	131.243.235	3043	131.243. 14	314
128. 3.188	102721	131.243.146	2797	131.243.144	308
131.243. 84	100839	131.243.202	2252	128. 3. 37	298
128. 3.208	100220	131.243.101	2023	128. 3. 63	281
131.243.107	97737	131.243.104	1966	131.243.105	236
131.243. 92	68918	131.243.201	1846	128. 3. 4	190
131.243. 26	40405	128. 3.189	1753	131.243.234	186
131.243.200	35482	131.243.203	1707	131.243. 13	183
128. 3. 78	34303	131.243. 88	1676	131.243.208	168
131.243. 10	29984	128. 3. 70	1673	128. 3. 38	157
128. 3.112	28361	131.243.106	1657	131.243.125	141
131.243. 89	27702	131.243. 90	1567	128. 3. 23	141
128. 3.209	27474	128. 3.191	1332	128. 3. 6	130
131.243. 87	24305	128. 3. 40	1215	131.243.126	100
131.243. 63	23127	128. 3.110	1127	128. 3. 39	90
131.243. 86	20314	128. 3.190	1006	131.243. 42	70
131.243. 85	20210	128. 3. 41	995	128. 3. 5	54
128. 3.211	19691	128. 3.193	982	131.243.181	50
131.243.145	18322	131.243.127	933	131.243. 40	45
131.243.103	13351	128. 3.210	931	128. 3.174	36
131.243. 60	8068	128. 3.242	877	128. 3. 36	30
131.243.170	6965	128. 3. 48	844	131.243. 6	27
131.243. 62	6655	131.243.147	790	128. 3.205	1
131.243.219	4967	131.243. 94	772		

131.243.140 Outgoing web traffic, incoming web, SMB, LDAP, 24409 UDP, 6894955 TCP, 5442 SYN

131.243.141 Outgoing web traffic, incoming web, 132894 UDP, 900243 TCP, 5265 SYN

131.243.142 Outgoing web traffic, incoming web, ssh, 15957 UDP, 739677 TCP, 4729 SYN

131.243.143 Outgoing web traffic, incoming web, 10248 UDP, 503297 TCP, 5523 SYN

- Port 21 2004-12-15 18:12–19:12

131.243.12 Out: 80:684, In: 80: 733, 515: 280, 47670 UDP, 308337 TCP, 1323 SYN

131.243.13 Out: 80:630, In: 80: 744, 515: 238, 16526 UDP, 329474 TCP, 1700 SYN

131.243.14 Out: 80:845, In: 80: 895, 6696 UDP, 104277 TCP, 1110 SYN

131.243.15 Out: 80: 413, In: 80:460, 515: 158, 10414 UDP, 963054 TCP, 879 SYN

- Port 12 2004-12-16 19:16 – 2004-12-16 20:16
 - 131.243.155: 408936

- Port 1 2004-12-16 21:17–22:17

128.3.44 Out: 80: 36, In: 80: 36, 1280 UDP, 33971 TCP, 53 SYN

128.3.45 Out: 80: 178, In 80: 178, 728 UDP, 81681 TCP, 240 SYN

128.3.46 Out: 80: 335, 631: 125, In: 80: 335, 631: 125, 9331 UDP, 35009 TCP, 558 SYN

128.3.47 Out: 80: 176, In: 80: 176, 832 UDP, 14649 TCP, 227 SYN

128.3.100 908 UDP

- Port 12 2005-01-06 22:25 – 2005-01-06 13:26

- 131.243.155: 48171

- Hundreds more

- Last trace

- 2005-01-08 04:28 – 2005-01-08 05:28

- port 16

- 421,021 packets

- Overlaps 128.3.47.0/24 and 128.3.100.0/24 with first trace, otherwise, all externally facing networks distinct from those above.

- 15,395 UDP packets (3–4%)

- 2312 multicast packets (\downarrow 1%)

- externally facing local nets:

131.243.235	406432	128.3.205	81	128.3. 60	8
128. 3.210	14065	128.3.219	81	128.3.101	8
128. 3.161	8514	128.3. 47	80	128.3. 87	8
128. 3.209	1081	128.3.241	80	128.3.175	8
128. 3.148	591	128.3.247	79	128.3.151	8
128. 3.250	527	128.3.165	78	128.3.117	8
128. 3. 71	419	128.3. 5	78	128.3. 10	8
128. 3.193	320	128.3. 81	77	128.3.192	8
128. 3.255	255	128.3.226	75	128.3. 40	6
128. 3. 88	167	128.3.243	72	128.3.231	6
128. 3. 0	158	128.3. 9	69	128.3.172	6
128. 3.181	152	128.3.100	68	128.3. 95	6
128. 3. 68	116	128.3.232	65	128.3. 78	6
128. 3.132	115	128.3.129	58	128.3.115	6
128. 3. 49	92	128.3. 58	58	128.3.150	6
128. 3.202	87	128.3.170	55	128.3.166	6
128. 3.238	83	128.3. 14	48	128.3.106	5
128. 3.154	82	128.3. 83	43	128.3.105	5
128. 3. 22	82	128.3.113	30	128.3.120	4
128. 3. 77	82	128.3.235	22	128.3.133	4

– best internal facing local nets:

131.243.140	341719	128. 3. 36	145	131.243.102	39
131.243.141	6993	131.243. 94	128	128. 3. 43	8
131.243. 10	2328	131.243. 85	125	131.243.103	8
131.243.143	323	131.243.219	116	131.243.208	8
131.243.142	245	128. 3. 19	69	131.243.170	2
128. 3.147	232	128. 3. 63	45	131.243.144	2
131.243.105	203	128. 3. 41	44		

- Other than the last trace, which takes place months after the first four, each trace has distinctly different externally facing nets, and while some external facing nets on one trace may be internal facing nets on another, the low packet rates are consistent with those being the other end of connections for nets that the port actually covers.
- Examined three additional traces: two for port 1 and one for port 2, just to compare their network coverage to the traces for port 1 and port 2, above.
 - First trace (port 1) perfectly consistent with externally facing nets, half of the internally facing nets show up again, with two “new” ones.
 - Second trace (port 2) same top, highly dominate externally facing net; two of the previously internally facing nets now externally facing; some but not all of the previously externally facing nets (a significant number of 131.243.X.0/24 nets) show up as internally facing; top internally facing net consistent; other internally facing nets come and go.

- Third trace (port 1) again, perfectly consistent with externally facing nets, and same sort of retention and change-over with internal nets.
- <http://bro-ids.org/enterprise-traces/hdr-traces05/>

I.8 DefCon CTF

I.8.1 Shmoo data

Available at <http://cctf.shmoo.com/data/>

Dataset: DefCon 8

Appears to be corrupted

Dataset: DefCon 10

- Orange
 - 2002-08-03 00:59 – 2002-08-04 19:56:21
 - 1,118,784 packets
 - 192.168.2.0 and 192.168.36.0 local nets

I.8.2 Dataset: DefCon 9

- <http://public.www.planetmirror.com/pub/cctf/defcon9/?fl=>
- First tracefile
 - 2,567,529 packets
 - 2001-07-13 – 2001-07-14 (7 hours)
 - Local nets:

10.255. 10	805064
10.255. 40	477070
123.123.123	369078
10.255. 0	276010

I.9 DSL1

- 2006-10-05 – 2006-10-07 (partial on vir)
- 2006-10-08 – 2006-12-28
- 2007-01-27 – 2007-04-25
- 2007-07-28 – 2007-08-22

I.10 Linode

- 2007-07-20 – 2007-11-09
 - break on 2007-07-30
 - 5,300,270 IP packets
- 2004-09-16 – 2005-02-13
 - 12,409,907 IP packets

I.11 DSL2

- Four minutes from testing

I.12 Data timeline

Start date	End date	Dataset	Intersects with
1999-02-23	Present	MAWI	Everything else
2001-07-13	2001-07-14	DefCon 9	
2001-08-29	2001-08-31	SIGCOMM2001	
2001-09-20	2002-01-10	Dartmouth	
2002-03-21	2002-06-09	Dartmouth	
2002-08-03	2002-08-04	DefCon 10	
2002-08-14	2002-08-14	Ames	
2003-01-15	2003-01-15	Ames	
2003-03-01	2003-03-06	SOTM27	
2003-04-24	2003-04-24	Ames	
2003-11-01	2004-02-27	Dartmouth	
2004-08-30	2004-09-03	SIGCOMM2004	
2004-09-16	2005-02-13	Linode	LBNL
2004-10-04	2004-10-04	LBNL	Linode
2004-12-15	2004-12-16	LBNL	Linode
2005-01-06	2005-01-07	LBNL	Linode
2005-03-09	2005-03-10	IETF2005	
2006-01-10	2006-01-11	DNS and KAIST&KOREN	
2006-06-12	2006-06-12	PDX	
2006-06-26	2006-06-26	PDX	
2006-07-10	2006-07-10	(2) PDX	
2006-07-20	2006-07-20	PDX	
2006-07-24	2006-07-24	PDX	
2006-10-05	2006-10-07	DSL1	
2006-10-08	2006-10-22	DSL1	
2006-11-06	2006-11-07	OSDI2006	
2007-01-08	2007-01-11	Ampath	
2007-01-27	2007-04-25	DSL1	
2007-07-20	Present	Linode	DSL1
2007-07-28	2007-08-22	DSL1	Linode

I.13 Data Constraints

Chapter 8 covers the pairs of traces that we use with a basecase-centric presentation. In order to construct that, we needed to look at our available data in a dataset-centric manner, enumerate the constraints we had to fill the needs of the basecases, and ultimately choose pairs of data. Part of this process was trying to minimize the number of traces we needed for the tests to optimize storage; this was accomplished by attempting to choose traces that would be usable in multiple trace-pairs where ever possible.

I.13.1 LBNL

- 1 Hour: Different datasets

Linode 128.3.45.0/24 on 2004-12-16 21:17 and 64.5.53.67/32 on 2004-12-16 21:17

Linode¹ 128.3.23.0/24 on 2005-01-07 23:26 and 64.5.53.67/32 on 2005-01-07 23:26

- 1 Hour: Different nets in same dataset

Different LBNL net 128.3.45.0/24 on 2004-12-16 21:17 and 128.3.46.0/24 on 2004-12-16 21:17

Similar LBNL net 131.243.12.0/24 on 2004-12-15 18:12 and 131.243.13.0/24 on 2004-12-15 18:12

Another similar LBNL net 128.3.45.0/24 on 2004-12-16 21:17 and 128.3.47.0/24 on 2004-12-16 21:17

- 1 Hour: Different times, same net

Mid-day/Off-hour 131.243.140.0/24 on 2004-12-15 11:09 and 131.243.140.0/24 on 2004-12-15 19:42

+1 week, Mid-day 128.3.23.0/24 on 2004-12-16 21:17 and 128.3.23.0/24 on 2005-01-07 23:26

Another +1 week, Mid-day 131.243.155.0/24 on 2004-12-16 21:17 and 131.243.155.0/24 on 2005-01-06 22:25

I.13.2 DefCon10

- 1 Day: Different datasets

Linode² 192.168.2.0/24 on 2002-08-03 12:00 and 64.5.53.67/32 on 2007-08-04 12:00

DSL1³ 192.168.2.0/24 on 2002-08-03 12:00 and 71.133.175.192/32 on 2007-08-04 12:00

I.13.3 Ames

- 1 Hour: Different nets in same dataset

Similar Ames net 237.24.63.0/24 on 2002-08-14 16:00 and 237.24.65.0/24 on 2002-08-14 16:00

¹Different month

²Same day (first Saturday in August), years apart

³Same day, (first Saturday in August) years apart

Another similar Ames net 239.65.134.0/24 on 2002-08-14 16:00 and 3.12.2.0/24 on 2002-08-14 16:00

Different Ames net 173.148.249.0/24 on 2002-08-14 16:00 and 3.3.31.0/24 on 2002-08-14 16:00

- 1 Hour: Different times, same net

+1 hour, Mid-day 237.24.63.0/24 on 2002-08-14 16:00 and 237.24.63.0/24 on 2002-08-14 17:00

Another +1 hour, Mid-day 239.65.134.0/24 on 2002-08-14 16:00 and 239.65.134.0/24 on 2002-08-14 17:00

I.13.4 Dartmouth

- 1 Day: Different datasets

Linode⁴ 190.84.172.0/24 on 2003-11-03 00:00 and 64.5.53.67/32 on 2007-11-05 00:00

DSL1⁵ 190.84.172.0/24 on 2004-01-28 00:00 and 69.110.74.139/32 on 2007-02-01 00:00

SOTM27⁶ 190.84.172.0/24 on 2003-11-03 00:00 and 172.16.134.0/24 on 2003-03-02 00:00

SIGCOMM2004⁷ 190.84.172.0/24 on 2004-01-27 14:00 and 98.93.251.0/24 on 2004-08-31 14:00

- 1 Day: Different nets in same dataset

Different Dartmouth net 190.84.172.0/24 on 2003-11-03 00:00 and 190.84.116.0/24 on 2003-11-03 00:00

Another different Dartmouth net 190.84.172.0/24 on 2003-11-03 00:00 and 190.84.224.0/24 on 2003-11-03 00:00

Similar Dartmouth net 190.84.172.0/24 on 2003-11-03 00:00 and 190.84.69.0/24 on 2003-11-03 00:00

Another similar Dartmouth net 190.84.172.0/24 on 2003-11-03 00:00 and 190.84.44.0/24 on 2003-11-03 00:00

⁴Same day (first Monday in November), years apart

⁵Same day (day after last Wednesday in January), years apart

⁶Months apart

⁷Months apart

- 1 Day: Different times, same net

+1 week 190.84.172.0/24 on 2003-11-03 00:00 and 190.84.172.0/24 on 2003-11-10 00:00

Another +1 week 190.84.69.0/24 on 2003-11-03 00:00 and 190.84.69.0/24 on 2003-11-10 00:00

- 1 Hour: Different datasets

Linode⁸ 190.84.172.0/24 on 2003-11-03 18:00 and 64.5.53.67/32 on 2007-11-05 18:00

DSL1⁹ 190.84.172.0/24 on 2004-01-29 18:00 and 69.110.74.139/32 on 2007-02-01 18:00

- 1 Hour: Different nets in same dataset

Different Dartmouth net 190.84.172.0/24 on 2003-11-03 18:00 and 190.84.116.0/24 on 2003-11-03 18:00

Similar Dartmouth net 190.84.172.0/24 on 2003-11-03 18:00 and 190.84.69.0/24 on 2003-11-03 18:00

Another similar Dartmouth net 190.84.172.0/24 on 2003-11-03 18:00 and 190.84.44.0/24 on 2003-11-03 18:00

- 1 Hour: Different times, same net

Mid-day/Off-hour 190.84.172.0/24 on 2003-11-03 10:00 and 190.84.172.0/24 on 2003-11-03 18:00

+1 week, Off-hour 190.84.172.0/24 on 2003-11-03 10:00 and 190.84.172.0/24 on 2003-11-10 10:00

Another +1 week, Off-hour 190.84.69.0/24 on 2003-11-03 10:00 and 190.84.69.0/24 on 2003-11-10 10:00

+1 week, Mid-day 190.84.172.0/24 on 2003-11-03 18:00 and 190.84.172.0/24 on 2003-11-10 18:00

Another +1 week, Mid-day 190.84.69.0/24 on 2003-11-03 18:00 and 190.84.69.0/24 on 2003-11-10 18:00

+1 hour, Off-hour 190.84.172.0/24 on 2003-11-03 09:00 and 190.84.172.0/24 on 2003-11-03 10:00

⁸Same time and day (first Monday in November), years apart

⁹Same time and day (day after last Wednesday in January), years apart

Another +1 hour, Off-hour 190.84.69.0/24 on 2003-11-03 09:00 and 190.84.69.0/24 on 2003-11-03 10:00

I.13.5 DSL1

- 1 Day: Different datasets

Dartmouth¹⁰ 190.84.172.0/24 on 2004-01-28 00:00 and 69.110.74.139/32 on 2007-02-01 00:00

SOTM27¹¹ 69.225.88.159/32 on 2007-03-02 00:00 and 172.16.134.0/24 on 2003-03-02 00:00

DefCon10¹² 192.168.2.0/24 on 2002-08-03 12:00 and 71.133.175.192/32 on 2007-08-04 12:00

Linode 69.110.79.251/32 on 2007-08-06 00:00 and 64.5.53.67/32 on 2007-08-06 00:00

- 1 Day: Different times, same net

+1 week 69.110.79.251/32 on 2007-08-06 00:00 and 69.110.79.251/32 on 2007-08-13 00:00

- 1 Hour: Different datasets

Dartmouth¹³ 190.84.172.0/24 on 2004-01-29 18:00 and 69.110.74.139/32 on 2007-02-01 18:00

SOTM27¹⁴ 69.225.88.159/32 on 2007-03-02 18:00 and 172.16.134.0/24 on 2003-03-02 18:00

Linode 69.110.79.251/32 on 2007-08-06 18:00 and 64.5.53.67/32 on 2007-08-06 18:00

Another Linode 69.110.79.251/32 on 2007-08-13 18:00 and 64.5.53.67/32 on 2007-08-13 18:00

- 1 Hour: Different times, same net

Mid-day/Off-hour 69.110.79.251/32 on 2007-08-06 10:00 and 69.110.79.251/32 on 2007-08-06 18:00

¹⁰Same day (day after last Wednesday in January), years apart

¹¹Same day, years apart

¹²Same day, (first Saturday in August) years apart

¹³Same time and day (day after last Wednesday in January), years apart

¹⁴Same time and day, years apart

+1 week, Off-hour 69.110.79.251/32 on 2007-08-06 10:00 and 69.110.79.251/32 on 2007-08-13 10:00

+1 week, Mid-day 69.110.79.251/32 on 2007-08-06 18:00 and 69.110.79.251/32 on 2007-08-13 18:00

+1 hour, Off-hour 69.110.79.251/32 on 2007-08-06 09:00 and 69.110.79.251/32 on 2007-08-06 10:00

+1 hour, Mid-day 69.110.79.251/32 on 2007-08-06 18:00 and 69.110.79.251/32 on 2007-08-06 19:00

I.13.6 Linode

- 1 Day: Different datasets

Dartmouth ¹⁵ 190.84.172.0/24 on 2003-11-03 00:00 and 64.5.53.67/32 on 2007-11-05 00:00

SOTM27 ¹⁶ 64.5.53.67/32 on 2008-03-02 00:00 and 172.16.134.0/24 on 2003-03-02 00:00

DefCon10 ¹⁷ 192.168.2.0/24 on 2002-08-03 12:00 and 64.5.53.67/32 on 2007-08-04 12:00

DSL1 69.110.79.251/32 on 2007-08-06 00:00 and 64.5.53.67/32 on 2007-08-06 00:00

- 1 Day: Different times, same net

+1 week 64.5.53.67/32 on 2007-08-06 00:00 and 64.5.53.67/32 on 2007-08-13 00:00

- 1 Hour: Different datasets

Dartmouth ¹⁸ 190.84.172.0/24 on 2003-11-03 18:00 and 64.5.53.67/32 on 2007-11-05 18:00

LBNL 128.3.45.0/24 on 2004-12-16 21:17 and 64.5.53.67/32 on 2004-12-16 21:17

LBNL ¹⁹ 128.3.23.0/24 on 2005-01-07 23:26 and 64.5.53.67/32 on 2005-01-07 23:26

SOTM27 ²⁰ 64.5.53.67/32 on 2008-03-02 18:00 and 172.16.134.0/24 on 2003-03-02 18:00

DSL1 69.110.79.251/32 on 2007-08-06 18:00 and 64.5.53.67/32 on 2007-08-06 18:00

Another DSL1 69.110.79.251/32 on 2007-08-13 18:00 and 64.5.53.67/32 on 2007-08-13 18:00

¹⁵Same day (first Monday in November), years apart

¹⁶Same day, years apart

¹⁷Same day (first Saturday in August), years apart

¹⁸Same time and day (first Monday in November), years apart

¹⁹Different month

²⁰Same time and day, years apart

- 1 Hour: Different times, same net

Mid-day/Off-hour 64.5.53.67/32 on 2007-08-06 10:00 and 64.5.53.67/32 on 2007-08-06

18:00

+1 week, Off-hour 64.5.53.67/32 on 2007-08-06 10:00 and 64.5.53.67/32 on 2007-08-13

10:00

+1 week, Mid-day 64.5.53.67/32 on 2007-08-06 18:00 and 64.5.53.67/32 on 2007-08-13

18:00

+1 hour, Off-hour 64.5.53.67/32 on 2007-08-06 09:00 and 64.5.53.67/32 on 2007-08-06

10:00

+1 hour, Mid-day 64.5.53.67/32 on 2007-08-06 18:00 and 64.5.53.67/32 on 2007-08-06

19:00

I.13.7 SOTM27

- 1 Day: Different datasets

Linode²¹ 64.5.53.67/32 on 2008-03-02 00:00 and 172.16.134.0/24 on 2003-03-02 00:00

DSL1²² 69.225.88.159/32 on 2007-03-02 00:00 and 172.16.134.0/24 on 2003-03-02 00:00

SIGCOMM2004 172.16.134.0/24 on 2003-03-04 14:00 and 98.93.251.0/24 on 2004-08-31

14:00

Dartmouth²³ 190.84.172.0/24 on 2003-11-03 00:00 and 172.16.134.0/24 on 2003-03-02

00:00

- 1 Day: Different times, same net

+3 days 172.16.134.0/24 on 2003-03-02 00:00 and 172.16.134.0/24 2003-03-05 00:00

- 1 Hour: Different datasets

Linode²⁴ 64.5.53.67/32 on 2008-03-02 18:00 and 172.16.134.0/24 on 2003-03-02 18:00

DSL1²⁵ 69.225.88.159/32 on 2007-03-02 18:00 and 172.16.134.0/24 on 2003-03-02 18:00

²¹Same day, years apart

²²Same day, years apart

²³Same day, months apart

²⁴Same time and day, years apart

²⁵Same time and day, years apart

SIGCOMM2004 172.16.134.0/24 on 2003-03-02 18:00 and 98.93.251.0/24 on 2004-08-31

19:00

- 1 Hour: Different times, same net

Mid-day/Off-hour 172.16.134.0/24 on 2003-03-02 10:00 and 172.16.134.0/24 on 2003-03-

02 18:00

+1 hour, Off-hour 172.16.134.0/24 on 2003-03-02 09:00 and 172.16.134.0/24 on 2003-03-02 10:00

+1 hour, Mid-day 172.16.134.0/24 on 2003-03-02 18:00 and 172.16.134.0/24 on 2003-03-02 19:00

+3 days, Off-hour 172.16.134.0/24 on 2003-03-02 10:00 and 172.16.134.0/24 on 2003-03-05 10:00

I.13.8 SIGCOMM2004

- 1 Day: Different datasets

SOTM27 172.16.134.0/24 on 2003-03-04 14:00 and 98.93.251.0/24 on 2004-08-31 14:00

Dartmouth²⁶ 190.84.172.0/24 on 2004-01-27 14:00 and 98.93.251.0/24 on 2004-08-31 14:00

- 1 Day: Different times, same net

+2 days 98.93.251.0/24 on 2004-08-31 14:00 and 98.93.251.0/24 on 2004-09-02 14:00

- 1 Hour: Different datasets

SOTM27 172.16.134.0/24 on 2003-03-02 18:00 and 98.93.251.0/24 on 2004-08-31 19:00

- 1 Hour: Different times, same net

Mid-day/Off-hour 98.93.251.0/24 on 2004-09-01 10:00²⁷ and 98.93.251.0/24 on 2004-08-31 19:00

²⁶Months apart

²⁷This trace is actually empty, but it was not a sensor outage: we see ARP traffic during this period

+1 hour, Off-hour 98.93.251.0/24 on 2004-09-01 09:00²⁸ and 98.93.251.0/24 on 2004-09-01 10:00

+1 hour, Mid-day 98.93.251.0/24 on 2004-08-31 19:00 and 98.93.251.0/24 on 2004-08-31 20:00

+2 days, Off-hour 98.93.251.0/24 on 2004-09-01 10:00 and 98.93.251.0/24 on 2004-09-03 10:00²⁹

I.14 Unused descriptions

Chapter 8 provides a description for each dataset that was used for the tests on real data. Some of the datasets ended up not being necessary, so we are capturing their descriptions here.

I.14.1 MAWI

Description: The MAWI Working Group Traffic Archive is part of the WIDE project and makes available daily samples of traffic from one or more of their Japan/US peering points.

Topology / Policy: Traffic from numerous networks of differing topologies and unknown policies are captured.

Limitations:

- While anonymization appears consistent within a single trace, it appears to be different from day to day
- Not sure we are capturing complete flow data (due to asynchronous routing)
- Traces are only 15 minutes in length
- No idea what the policies are of the networks captured by the trace

²⁸This trace is actually empty, but it was not a sensor outage: we see ARP traffic during this period

²⁹This trace is actually empty, but it was not a sensor outage: we see ARP traffic during this period

Usefulness:

- Daily data going back to 23 February 1999 – much larger time window than any other dataset
- Could extract one subnet to compare against a trace from another dataset (made at the same time) on the presumption that the policies are significantly different, perhaps with human verification
- Could use by extracting one subnet, splitting extraction into two 7.5 minute sections and comparing those
- Could extract two different networks from same trace based on the presumption that the policies are significantly different if the traffic volume is significantly different and compare those
- Could take one subnet from one trace and compare it to each of the subnets in an adjacent trace and see if we can identify which network is the same, problem is this approach does not have any ground truth
- Similarly, could split the trace into two 7.5 minute traces, compare one subnet in the first half to each subnet in the second half and see if it identifies itself as the most similar

I.14.2 Ampath OC-12

Description: Ampath is a peering point in Florida. Anonymized packet header traces were made of the traffic on one of its links over two days for the Day in the Life (DitL) of the Internet project in 2007.

Topology / Policy: Traffic from numerous networks of differing topologies and unknown policies are captured.

Limitations:

- Not sure we are capturing complete flow data (due to asynchronous routing)
- No idea what the policies are of the networks captured by the trace
- Does not overlap with any other datasets

Usefulness:

- Could use by extracting one subnet, splitting extraction into two halves and comparing those
- Could extract two different networks based on the presumption that the policies are significantly different if the traffic volume is significantly different (perhaps with human verification) and compare those
- Could extract one network from this dataset and a network from another peering point dataset – such as the Ames Exchange data – which we can verify by hand to be different, and compare those

Appendix J

Real Data Details

This appendix contains additional tables and figures for chapter 8 which were not directly necessary for the accompanying discussion in that chapter.

Table J.1: Traces used for each training pair for each basecase, along with the intended goal and calculated mean similarity of all normalized characteristics.

Basecase	Pair	Trace1	Trace2	Goal	Mean
1	1	ten 64.5.53.67 20070806-1800	ten 64.5.53.67 20070806-1900	0.75 - 1.0	0.57309
1	2	sotm27 172.16.134 20030302-1800	sotm27 172.16.134 20030302-1900	0.75 - 1.0	0.00000
1	3	ames 237.24.63 20020814-1600	ames 237.24.63 20020814-1700	0.75 - 1.0	0.50336
2	1	ten 64.5.53.67 20070806-0900	ten 64.5.53.67 20070806-1000	0.75 - 1.0	0.61317
2	2	sotm27 172.16.134	sotm27 172.16.134	0.75 - 1.0	0.00000
Continued on next page					

Table J.1 – continued from previous page

Basecase	Pair	Trace1	Trace2	Goal	Mean
		20030302-0900	20030302-1000		
2	3	dartmouth 190.84.172 20031103-0900	dartmouth 190.84.172 20031103-1000	0.75 - 1.0	0.80514
3	1	ten 64.5.53.67 20070806-1800	ten 64.5.53.67 20070813-1800	0.75 - 1.0	0.58410
3	2	dartmouth 190.84.172 20031103-1800	dartmouth 190.84.172 20031110-1800	0.75 - 1.0	0.56727
3	3	dartmouth 190.84.69 20031103-1800	dartmouth 190.84.69 20031110-1800	0.75 - 1.0	0.74757
4	1	ten 64.5.53.67 20070806-1000	ten 64.5.53.67 20070813-1000	0.75 - 1.0	0.71072
4	2	dartmouth 190.84.172 20031103-1000	dartmouth 190.84.172 20031110-1000	0.75 - 1.0	0.55553
4	3	sotm27 172.16.134 20030302-1000	sotm27 172.16.134 20030305-1000	0.75 - 1.0	0.43909
5	1	ten 64.5.53.67 20070806-1000	ten 64.5.53.67 20070806-1800	Lower	0.54999
5	2	dartmouth 190.84.172 20031103-1000	dartmouth 190.84.172 20031103-1800	Lower	0.44249
5	3	sotm27 172.16.134 20030302-1000	sotm27 172.16.134 20030302-1800	Lower	0.72138
Continued on next page					

Table J.1 – continued from previous page

Basecase	Pair	Trace1	Trace2	Goal	Mean
6	1	ames 237.24.63 20020814-1600	ames 237.24.65 20020814-1600	0.5 - 0.75	0.75809
6	2	dartmouth 190.84.172 20031103-1800	dartmouth 190.84.69 20031103-1800	0.5 - 0.75	0.44971
6	3	dartmouth 190.84.172 20031103-1800	dartmouth 190.84.44 20031103-1800	0.5 - 0.75	0.58414
7	1	dsl1 69.110.79.251 20070806-1800	ten 64.5.53.67 20070806-1800	0.25 - 0.5	0.52037
7	2	dartmouth 190.84.116 20031103-1800	dartmouth 190.84.172 20031103-1800	0.25 - 0.5	0.60158
7	3	sigcomm2004 98.93.251 20040831-1900	sotm27 172.16.134 20030302-1800	0.25 - 0.5	0.40438
8	1	dartmouth 190.84.172 20031103-1800	ten 64.5.53.67 20071105-1800	0.0 - 0.25	0.45486
8	2	lbl 128.3.45 20041216-2117	ten 64.5.53.67 20041216-2117	0.0 - 0.25	0.52295
8	3	sotm27 172.16.134 20030302-1800	ten 64.5.53.67 20080302-1800	0.0 - 0.25	0.46846
9	1	ten 64.5.53.67 20070806-0000	ten 64.5.53.67 20070813-0000	0.75 - 1.0	0.72142
		dartmouth	dartmouth		
Continued on next page					

Table J.1 – continued from previous page

Basecase	Pair	Trace1	Trace2	Goal	Mean
9	2	190.84.172 20031103-0000	190.84.172 20031110-0000	0.75 - 1.0	0.58145
9	3	sotm27 172.16.134 20030302-0000	sotm27 172.16.134 20030305-0000	0.75 - 1.0	0.41186
10	1	dartmouth 190.84.172 20031103-0000	dartmouth 190.84.69 20031103-0000	0.5 - 0.75	0.49667
11	1	dsl1 69.110.79.251 20070806-0000	ten 64.5.53.67 20070806-0000	0.25 - 0.5	0.53287
11	2	dartmouth 190.84.116 20031103-0000	dartmouth 190.84.172 20031103-0000	0.25 - 0.5	0.59175
11	3	dartmouth 190.84.172 20031103-0000	sotm27 172.16.134 20030302-0000	0.25 - 0.5	0.52302
12	1	dartmouth 190.84.172 20031103-0000	ten 64.5.53.67 20071105-0000	0.0 - 0.25	0.45696
12	2	sotm27 172.16.134 20030302-0000	ten 64.5.53.67 20080302-0000	0.0 - 0.25	0.46348
12	3	dc10 192.168.2 20020803-1200	ten 64.5.53.67 20070804-1200	0.0 - 0.25	0.46809

Table J.2: Scaled, linear, and linear with zeros goals for all training pairs of real data basecases.

Basecase	Pair	Scaled Goal	Linear Goal	Linear w/ zeros
1	1	0.835	0.875	0.875
1	2	0.000	0.875	0.000
1	3	0.800	0.875	0.875
2	1	0.855	0.875	0.875
2	2	0.000	0.875	0.000
2	3	0.950	0.875	0.875
3	1	0.840	0.875	0.875
3	2	0.830	0.875	0.875
3	3	0.920	0.875	0.875
4	1	0.900	0.875	0.875
4	2	0.825	0.875	0.875
4	3	0.765	0.875	0.875
5	1	0.700	0.675	0.675
5	2	0.625	0.675	0.675
5	3	0.725	0.675	0.675
6	1	0.575	0.575	0.575
6	2	0.525	0.575	0.575
6	3	0.550	0.575	0.575
7	1	0.390	0.375	0.375
7	2	0.500	0.375	0.375
7	3	0.250	0.375	0.375
8	1	0.158	0.125	0.125
8	2	0.226	0.125	0.125
8	3	0.171	0.125	0.125
9	1	0.910	0.875	0.875
9	2	0.838	0.875	0.875
9	3	0.750	0.875	0.875
10	1	0.625	0.625	0.625
11	1	0.400	0.375	0.375

Continued on next page

Table J.2 – continued from previous page

Basecase	Pair	Scaled Goal	Linear Goal	Linear w/ zeros
11	2	0.490	0.375	0.375
11	3	0.350	0.375	0.375
12	1	0.160	0.125	0.125
12	2	0.166	0.125	0.125
12	3	0.171	0.125	0.125

Table J.3: Weights calculated by linear regression with the scaled goals.

Metric	Weight
y-intercept	0.875
Connection Destination IP nonKeyedSortedContinuous	-0.1925
Connections time rate sortedContinuous	-0.1963
Bytes in count	0.1185
Connection Priv packet rate sortedContinuous	0.2151
Priv packets time rate sortedContinuous	-0.2287
Connection min orderedContinuous	-0.134
Packet Destination IP nonKeyedSortedContinuous	0.2808
Number other errors orderedContinuous	0.1915
Packet TTL discrete	0.231
Connections out count	-0.2127
Resend rate sortedContinuous	-0.2141
Connection sec orderedContinuous	-0.1092
Connection Source port discrete	-0.0199
Number of packets in orderedContinuous	-0.1496
Packet Service discrete	-0.0716
Number connection errors orderedContinuous	0.1001
FINs connection time rate sortedContinuous	-0.5003
Unpriv packets time rate sortedContinuous	0.2238
Continued on next page	

Table J.3 – continued from previous page

Metric	Weight
Half-open connection rate ratio	-0.1756
RSTs connection time rate sortedContinuous	0.1971
Connection Source IP nonKeyedSortedContinuous	-0.0958
Priv connections connection time rate sortedContinuous	-0.2282
Bytes weekday orderedContinuous	-0.0748
Number data packets rate sortedContinuous	0.094
Connection weekday orderedContinuous	-0.153
Connection packet rate sortedContinuous	-0.129
Establishment errors connection time rate sortedContinuous	-0.1039
Packet Source port discrete	-0.0273
Connection Establishment errors rate sortedContinuous	0.1781
InterPacket delta sortedContinuous	-0.2662
Unpriv packets packet rate sortedContinuous	0.092
Connection RSTs rate sortedContinuous	0.1534
Connection Source port orderedContinuous	0.0113
Connection Service discrete	-0.0383
SYN-ACK rate ratio	-0.1364
InterConnection delta sortedContinuous	0.2474
Number bytes transferred in orderedContinuous	0.1839
Ave duration over last m connections sortedContinuous	-0.1978
Packets in count	0.1625
Packets in last w secs orderedContinuous	-0.0309
Unpriv connections connection time rate sortedContinuous	0.3721
Packet weekday orderedContinuous	-0.1025
Connection Unpriv packet rate sortedContinuous	-0.1526
Connection Unpriv connections rate sortedContinuous	0.1462
Priv packets packet rate sortedContinuous	-0.1692
Ave duration over last w secs sortedContinuous	0.2015
Duration sortedContinuous	0.1128
Connection PSH rate sortedContinuous	0.2164
Number data bytes transferred in orderedContinuous	-0.1259
Continued on next page	

Table J.3 – continued from previous page

Metric	Weight
Unpriv packets unpriv connection time rate sortedContinuous	0.1393
Connection FINs rate sortedContinuous	0.2143

Table J.4: Both our calculated scaled similarity (rounded to five significant digits) and Weka’s scaled similarity (rounded to three significant digits) for all test pairs of real data, using the scaled training goals.

Basecase	Pair	Our calculation	Weka calculation
1	1	0.87668	0.877
1	2	0.78035	0.78
1	3	0.59421	0.539
2	1	0.70184	0.702
2	2	0.00000	0.552
2	3	0.99457	0.995
3	1	0.53818	0.538
3	2	0.69100	0.691
3	3	0.43407	0.434
4	1	0.71139	0.711
4	2	0.46651	0.278
4	3	0.00000	0.552
5	1	0.53657	0.537
5	2	0.36291	0.363
5	3	0.00000	0
6	1	0.61716	0.561
6	2	0.50028	0.5
6	3	0.61807	0.618
7	1	0.23892	0.239
7	2	0.39955	0.4
Continued on next page			

Table J.4 – continued from previous page

Basecase	Pair	Our calculation	Weka calculation
7	3	0.41008	0.41
8	1	0.51419	0.645
8	2	0.53064	0.452
8	3	0.53471	0.505
9	1	0.69798	0.698
9	2	0.61866	0.619
9	3	0.77030	0.77
10	1	0.60454	0.604
11	1	0.10851	0.109
11	2	0.32791	0.328
11	3	0.55573	0.556
12	1	0.21439	0.214
12	2	0.40154	-0.032
12	3	0.32500	0.325

Appendix K

Real data distributions

With all of the traces of real data we are using, plotting the distributions for all the metrics in all those distributions takes a lot of space. Rather than clutter the main text with these tables and plots, we present them in this appendix.

Table K.1: Packets in counts for all traces.

Trace	Packets in
ames-173.148.249-20020814-1600	77994
ames-237.24.63-20020814-1600	11522
ames-237.24.63-20020814-1700	3884
ames-237.24.65-20020814-1600	13904
ames-239.65.134-20020814-1600	17079
ames-239.65.134-20020814-1700	14691
ames-3.12.2-20020814-1600	13162
ames-3.3.31-20020814-1600	0
dartmouth-190.84.116-20031103-0000	2536040
dartmouth-190.84.116-20031103-1800	16185
dartmouth-190.84.172-20031103-0000	817855
dartmouth-190.84.172-20031103-0900	1970
dartmouth-190.84.172-20031103-1000	1841
dartmouth-190.84.172-20031103-1800	77018
dartmouth-190.84.172-20031110-0000	5794299
Continued on next page	

Table K.1 – continued from previous page

Trace	packets in
dartmouth-190.84.172-20031110-1000	100557
dartmouth-190.84.172-20031110-1800	711541
dartmouth-190.84.172-20040127-1400	1147105
dartmouth-190.84.172-20040128-0000	464136
dartmouth-190.84.172-20040129-1800	158066
dartmouth-190.84.224-20031103-0000	391885
dartmouth-190.84.44-20031103-0000	3089808
dartmouth-190.84.44-20031103-1800	184267
dartmouth-190.84.69-20031103-0000	3448125
dartmouth-190.84.69-20031103-0900	728
dartmouth-190.84.69-20031103-1000	831
dartmouth-190.84.69-20031103-1800	241134
dartmouth-190.84.69-20031110-0000	6025393
dartmouth-190.84.69-20031110-1000	9488
dartmouth-190.84.69-20031110-1800	174605
dc10-192.168.2-20020803-1200	39027
dsl1-69.110.74.139-20070201-0000	50651
dsl1-69.110.74.139-20070201-1800	404
dsl1-69.110.79.251-20070806-0000	42925
dsl1-69.110.79.251-20070806-0900	902
dsl1-69.110.79.251-20070806-1000	2565
dsl1-69.110.79.251-20070806-1800	2257
dsl1-69.110.79.251-20070806-1900	1022
dsl1-69.110.79.251-20070813-0000	217093
dsl1-69.110.79.251-20070813-1000	3300
dsl1-69.110.79.251-20070813-1800	1203
dsl1-69.225.88.159-20070302-0000	545005
dsl1-69.225.88.159-20070302-1800	167
dsl1-71.133.175.192-20070804-1200	19737
lbl-128.3.23-20041216-2117	2267174
lbl-128.3.23-20050107-2326	734084
Continued on next page	

Table K.1 – continued from previous page

Trace	packets in
lbl-128.3.45-20041216-2117	47873
lbl-128.3.46-20041216-2117	23751
lbl-128.3.47-20041216-2117	7946
lbl-131.243.12-20041215-1812	209058
lbl-131.243.13-20041215-1812	206573
lbl-131.243.140-20041215-1109	80482
lbl-131.243.140-20041215-1942	2530336
lbl-131.243.155-20041216-2117	252915
lbl-131.243.155-20050106-2225	24808
sigcomm2004-98.93.251-20040831-1400	804968
sigcomm2004-98.93.251-20040831-1900	73593
sigcomm2004-98.93.251-20040831-2000	92023
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0
sigcomm2004-98.93.251-20040902-1400	949322
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	80
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	6
sotm27-172.16.134-20030302-1800	5
sotm27-172.16.134-20030302-1900	0
sotm27-172.16.134-20030304-1400	17190
sotm27-172.16.134-20030305-0000	17178
sotm27-172.16.134-20030305-1000	3436
ten-64.5.53.67-20041216-2117	109
ten-64.5.53.67-20050107-2326	311
ten-64.5.53.67-20070804-1200	6368
ten-64.5.53.67-20070806-0000	39396
ten-64.5.53.67-20070806-0900	112
ten-64.5.53.67-20070806-1000	376
ten-64.5.53.67-20070806-1800	9553

Continued on next page

Table K.1 – continued from previous page

Trace	packets in
ten-64.5.53.67-20070806-1900	1225
ten-64.5.53.67-20070813-0000	22305
ten-64.5.53.67-20070813-1000	680
ten-64.5.53.67-20070813-1800	753
ten-64.5.53.67-20071105-0000	13628
ten-64.5.53.67-20071105-1800	580
ten-64.5.53.67-20080302-0000	22150
ten-64.5.53.67-20080302-1800	927

Table K.2: Packets out counts for all traces.

Trace	packets out
ames-173.148.249-20020814-1600	282
ames-237.24.63-20020814-1600	0
ames-237.24.63-20020814-1700	0
ames-237.24.65-20020814-1600	0
ames-239.65.134-20020814-1600	0
ames-239.65.134-20020814-1700	0
ames-3.12.2-20020814-1600	0
ames-3.3.31-20020814-1600	167828
dartmouth-190.84.116-20031103-0000	2423842
dartmouth-190.84.116-20031103-1800	14479
dartmouth-190.84.172-20031103-0000	727884
dartmouth-190.84.172-20031103-0900	2444
dartmouth-190.84.172-20031103-1000	2323
dartmouth-190.84.172-20031103-1800	61517
dartmouth-190.84.172-20031110-0000	5494251
dartmouth-190.84.172-20031110-1000	32264
dartmouth-190.84.172-20031110-1800	341587
Continued on next page	

Table K.2 – continued from previous page

Trace	Packets out
dartmouth-190.84.172-20040127-1400	1032261
dartmouth-190.84.172-20040128-0000	335621
dartmouth-190.84.172-20040129-1800	109343
dartmouth-190.84.224-20031103-0000	327493
dartmouth-190.84.44-20031103-0000	2710469
dartmouth-190.84.44-20031103-1800	159235
dartmouth-190.84.69-20031103-0000	19350731
dartmouth-190.84.69-20031103-0900	350
dartmouth-190.84.69-20031103-1000	449
dartmouth-190.84.69-20031103-1800	1531093
dartmouth-190.84.69-20031110-0000	23066484
dartmouth-190.84.69-20031110-1000	479891
dartmouth-190.84.69-20031110-1800	1059254
dc10-192.168.2-20020803-1200	50931
dsl1-69.110.74.139-20070201-0000	30064
dsl1-69.110.74.139-20070201-1800	242
dsl1-69.110.79.251-20070806-0000	33748
dsl1-69.110.79.251-20070806-0900	731
dsl1-69.110.79.251-20070806-1000	2343
dsl1-69.110.79.251-20070806-1800	1498
dsl1-69.110.79.251-20070806-1900	704
dsl1-69.110.79.251-20070813-0000	137396
dsl1-69.110.79.251-20070813-1000	3203
dsl1-69.110.79.251-20070813-1800	902
dsl1-69.225.88.159-20070302-0000	157678
dsl1-69.225.88.159-20070302-1800	29
dsl1-71.133.175.192-20070804-1200	16734
lbl-128.3.23-20041216-2117	1739310
lbl-128.3.23-20050107-2326	699543
lbl-128.3.45-20041216-2117	35246
lbl-128.3.46-20041216-2117	17331
Continued on next page	

Table K.2 – continued from previous page

Trace	Packets out
lbl-128.3.47-20041216-2117	8228
lbl-131.243.12-20041215-1812	149174
lbl-131.243.13-20041215-1812	140561
lbl-131.243.140-20041215-1109	126757
lbl-131.243.140-20041215-1942	4394118
lbl-131.243.155-20041216-2117	154911
lbl-131.243.155-20050106-2225	21375
sigcomm2004-98.93.251-20040831-1400	784386
sigcomm2004-98.93.251-20040831-1900	65118
sigcomm2004-98.93.251-20040831-2000	89520
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0
sigcomm2004-98.93.251-20040902-1400	811410
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	60
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	4
sotm27-172.16.134-20030302-1800	4
sotm27-172.16.134-20030302-1900	0
sotm27-172.16.134-20030304-1400	14841
sotm27-172.16.134-20030305-0000	14825
sotm27-172.16.134-20030305-1000	3312
ten-64.5.53.67-20041216-2117	92
ten-64.5.53.67-20050107-2326	313
ten-64.5.53.67-20070804-1200	6187
ten-64.5.53.67-20070806-0000	30645
ten-64.5.53.67-20070806-0900	93
ten-64.5.53.67-20070806-1000	385
ten-64.5.53.67-20070806-1800	11461
ten-64.5.53.67-20070806-1900	959
ten-64.5.53.67-20070813-0000	15777
Continued on next page	

Table K.2 – continued from previous page

Trace	Packets out
ten-64.5.53.67-20070813-1000	672
ten-64.5.53.67-20070813-1800	726
ten-64.5.53.67-20071105-0000	13060
ten-64.5.53.67-20071105-1800	550
ten-64.5.53.67-20080302-0000	21396
ten-64.5.53.67-20080302-1800	822

Table K.3: Connections in counts for all traces.

Trace	Connections in
ames-173.148.249-20020814-1600	339
ames-237.24.63-20020814-1600	86
ames-237.24.63-20020814-1700	69
ames-237.24.65-20020814-1600	132
ames-239.65.134-20020814-1600	145
ames-239.65.134-20020814-1700	58
ames-3.12.2-20020814-1600	87
ames-3.3.31-20020814-1600	0
dartmouth-190.84.116-20031103-0000	76806
dartmouth-190.84.116-20031103-1800	3609
dartmouth-190.84.172-20031103-0000	184870
dartmouth-190.84.172-20031103-0900	632
dartmouth-190.84.172-20031103-1000	575
dartmouth-190.84.172-20031103-1800	9968
dartmouth-190.84.172-20031110-0000	69434
dartmouth-190.84.172-20031110-1000	683
dartmouth-190.84.172-20031110-1800	4617
dartmouth-190.84.172-20040127-1400	30565
dartmouth-190.84.172-20040128-0000	160668
Continued on next page	

Table K.3 – continued from previous page

Trace	Connections in
dartmouth-190.84.172-20040129-1800	6825
dartmouth-190.84.224-20031103-0000	157744
dartmouth-190.84.44-20031103-0000	281670
dartmouth-190.84.44-20031103-1800	39247
dartmouth-190.84.69-20031103-0000	668160
dartmouth-190.84.69-20031103-0900	513
dartmouth-190.84.69-20031103-1000	637
dartmouth-190.84.69-20031103-1800	37207
dartmouth-190.84.69-20031110-0000	491063
dartmouth-190.84.69-20031110-1000	7539
dartmouth-190.84.69-20031110-1800	31464
dc10-192.168.2-20020803-1200	567
dsl1-69.110.74.139-20070201-0000	5281
dsl1-69.110.74.139-20070201-1800	217
dsl1-69.110.79.251-20070806-0000	6910
dsl1-69.110.79.251-20070806-0900	306
dsl1-69.110.79.251-20070806-1000	391
dsl1-69.110.79.251-20070806-1800	354
dsl1-69.110.79.251-20070806-1900	346
dsl1-69.110.79.251-20070813-0000	8288
dsl1-69.110.79.251-20070813-1000	448
dsl1-69.110.79.251-20070813-1800	289
dsl1-69.225.88.159-20070302-0000	476154
dsl1-69.225.88.159-20070302-1800	166
dsl1-71.133.175.192-20070804-1200	5841
lbl-128.3.23-20041216-2117	117925
lbl-128.3.23-20050107-2326	26740
lbl-128.3.45-20041216-2117	611
lbl-128.3.46-20041216-2117	4177
lbl-128.3.47-20041216-2117	984
lbl-131.243.12-20041215-1812	24419
Continued on next page	

Table K.3 – continued from previous page

Trace	Connections in
lbl-131.243.13-20041215-1812	9244
lbl-131.243.140-20041215-1109	10998
lbl-131.243.140-20041215-1942	15016
lbl-131.243.155-20041216-2117	3943
lbl-131.243.155-20050106-2225	3870
sigcomm2004-98.93.251-20040831-1400	181339
sigcomm2004-98.93.251-20040831-1900	13668
sigcomm2004-98.93.251-20040831-2000	26926
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0
sigcomm2004-98.93.251-20040902-1400	199233
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	33
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	3
sotm27-172.16.134-20030302-1800	2
sotm27-172.16.134-20030302-1900	0
sotm27-172.16.134-20030304-1400	1412
sotm27-172.16.134-20030305-0000	1414
sotm27-172.16.134-20030305-1000	965
ten-64.5.53.67-20041216-2117	28
ten-64.5.53.67-20050107-2326	25
ten-64.5.53.67-20070804-1200	3280
ten-64.5.53.67-20070806-0000	4798
ten-64.5.53.67-20070806-0900	69
ten-64.5.53.67-20070806-1000	178
ten-64.5.53.67-20070806-1800	135
ten-64.5.53.67-20070806-1900	384
ten-64.5.53.67-20070813-0000	5301
ten-64.5.53.67-20070813-1000	380
ten-64.5.53.67-20070813-1800	425
Continued on next page	

Table K.3 – continued from previous page

Trace	Connections in
ten-64.5.53.67-20071105-0000	5170
ten-64.5.53.67-20071105-1800	297
ten-64.5.53.67-20080302-0000	9578
ten-64.5.53.67-20080302-1800	347

Table K.4: Connections out counts for all traces.

Trace	Connections out
ames-173.148.249-20020814-1600	30
ames-237.24.63-20020814-1600	0
ames-237.24.63-20020814-1700	0
ames-237.24.65-20020814-1600	0
ames-239.65.134-20020814-1600	0
ames-239.65.134-20020814-1700	0
ames-3.12.2-20020814-1600	0
ames-3.3.31-20020814-1600	706
dartmouth-190.84.116-20031103-0000	109826
dartmouth-190.84.116-20031103-1800	5555
dartmouth-190.84.172-20031103-0000	157167
dartmouth-190.84.172-20031103-0900	359
dartmouth-190.84.172-20031103-1000	314
dartmouth-190.84.172-20031103-1800	7570
dartmouth-190.84.172-20031110-0000	675741
dartmouth-190.84.172-20031110-1000	655
dartmouth-190.84.172-20031110-1800	2815
dartmouth-190.84.172-20040127-1400	152154
dartmouth-190.84.172-20040128-0000	167405
dartmouth-190.84.172-20040129-1800	9275
dartmouth-190.84.224-20031103-0000	204174
Continued on next page	

Table K.4 – continued from previous page

Trace	Connections out
dartmouth-190.84.44-20031103-0000	432247
dartmouth-190.84.44-20031103-1800	36190
dartmouth-190.84.69-20031103-0000	11639926
dartmouth-190.84.69-20031103-0900	290
dartmouth-190.84.69-20031103-1000	416
dartmouth-190.84.69-20031103-1800	929309
dartmouth-190.84.69-20031110-0000	12915508
dartmouth-190.84.69-20031110-1000	355601
dartmouth-190.84.69-20031110-1800	543853
dc10-192.168.2-20020803-1200	22533
dsl1-69.110.74.139-20070201-0000	3515
dsl1-69.110.74.139-20070201-1800	96
dsl1-69.110.79.251-20070806-0000	5069
dsl1-69.110.79.251-20070806-0900	157
dsl1-69.110.79.251-20070806-1000	396
dsl1-69.110.79.251-20070806-1800	174
dsl1-69.110.79.251-20070806-1900	141
dsl1-69.110.79.251-20070813-0000	8307
dsl1-69.110.79.251-20070813-1000	528
dsl1-69.110.79.251-20070813-1800	161
dsl1-69.225.88.159-20070302-0000	115746
dsl1-69.225.88.159-20070302-1800	29
dsl1-71.133.175.192-20070804-1200	3618
lbl-128.3.23-20041216-2117	117404
lbl-128.3.23-20050107-2326	30604
lbl-128.3.45-20041216-2117	835
lbl-128.3.46-20041216-2117	1652
lbl-128.3.47-20041216-2117	536
lbl-131.243.12-20041215-1812	25601
lbl-131.243.13-20041215-1812	9365
lbl-131.243.140-20041215-1109	10050
Continued on next page	

Table K.4 – continued from previous page

Trace	Connections out
lbl-131.243.140-20041215-1942	17578
lbl-131.243.155-20041216-2117	4908
lbl-131.243.155-20050106-2225	5157
sigcomm2004-98.93.251-20040831-1400	231070
sigcomm2004-98.93.251-20040831-1900	13659
sigcomm2004-98.93.251-20040831-2000	29529
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0
sigcomm2004-98.93.251-20040902-1400	240996
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	13
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	1
sotm27-172.16.134-20030302-1800	1
sotm27-172.16.134-20030302-1900	0
sotm27-172.16.134-20030304-1400	77
sotm27-172.16.134-20030305-0000	74
sotm27-172.16.134-20030305-1000	2
ten-64.5.53.67-20041216-2117	10
ten-64.5.53.67-20050107-2326	14
ten-64.5.53.67-20070804-1200	2969
ten-64.5.53.67-20070806-0000	4438
ten-64.5.53.67-20070806-0900	51
ten-64.5.53.67-20070806-1000	171
ten-64.5.53.67-20070806-1800	120
ten-64.5.53.67-20070806-1900	363
ten-64.5.53.67-20070813-0000	4894
ten-64.5.53.67-20070813-1000	358
ten-64.5.53.67-20070813-1800	406
ten-64.5.53.67-20071105-0000	4727
ten-64.5.53.67-20071105-1800	282
Continued on next page	

Table K.4 – continued from previous page

Trace	Connections out
ten-64.5.53.67-20080302-0000	8942
ten-64.5.53.67-20080302-1800	321

Table K.5: Bytes in counts for all traces.

Trace	Bytes in
ames-173.148.249-20020814-1600	75027220
ames-237.24.63-20020814-1600	4089671
ames-237.24.63-20020814-1700	485943
ames-237.24.65-20020814-1600	2654132
ames-239.65.134-20020814-1600	17907088
ames-239.65.134-20020814-1700	16471058
ames-3.12.2-20020814-1600	12343683
ames-3.3.31-20020814-1600	0
dartmouth-190.84.116-20031103-0000	1167955142
dartmouth-190.84.116-20031103-1800	9032621
dartmouth-190.84.172-20031103-0000	526646791
dartmouth-190.84.172-20031103-0900	227407
dartmouth-190.84.172-20031103-1000	198123
dartmouth-190.84.172-20031103-1800	63488676
dartmouth-190.84.172-20031110-0000	5045694672
dartmouth-190.84.172-20031110-1000	125179396
dartmouth-190.84.172-20031110-1800	802494763
dartmouth-190.84.172-20040127-1400	548238607
dartmouth-190.84.172-20040128-0000	412583602
dartmouth-190.84.172-20040129-1800	194665097
dartmouth-190.84.224-20031103-0000	242638945
dartmouth-190.84.44-20031103-0000	2590359240
dartmouth-190.84.44-20031103-1800	143280688
Continued on next page	

Table K.5 – continued from previous page

Trace	Bytes in
dartmouth-190.84.69-20031103-0000	2591121837
dartmouth-190.84.69-20031103-0900	77839
dartmouth-190.84.69-20031103-1000	83469
dartmouth-190.84.69-20031103-1800	205240833
dartmouth-190.84.69-20031110-0000	6673186820
dartmouth-190.84.69-20031110-1000	952249
dartmouth-190.84.69-20031110-1800	145270614
dc10-192.168.2-20020803-1200	5464793
dsl1-69.110.74.139-20070201-0000	45940033
dsl1-69.110.74.139-20070201-1800	138495
dsl1-69.110.79.251-20070806-0000	29433678
dsl1-69.110.79.251-20070806-0900	261794
dsl1-69.110.79.251-20070806-1000	1560711
dsl1-69.110.79.251-20070806-1800	1914771
dsl1-69.110.79.251-20070806-1900	296952
dsl1-69.110.79.251-20070813-0000	257811311
dsl1-69.110.79.251-20070813-1000	1618505
dsl1-69.110.79.251-20070813-1800	738220
dsl1-69.225.88.159-20070302-0000	173577073
dsl1-69.225.88.159-20070302-1800	51820
dsl1-71.133.175.192-20070804-1200	5393828
lbl-128.3.23-20041216-2117	3247654741
lbl-128.3.23-20050107-2326	182691920
lbl-128.3.45-20041216-2117	49295035
lbl-128.3.46-20041216-2117	19709834
lbl-128.3.47-20041216-2117	2450685
lbl-131.243.12-20041215-1812	236253074
lbl-131.243.13-20041215-1812	222230698
lbl-131.243.140-20041215-1109	11116883
lbl-131.243.140-20041215-1942	728793449
lbl-131.243.155-20041216-2117	90627132
Continued on next page	

Table K.5 – continued from previous page

Trace	Bytes in
lbl-131.243.155-20050106-2225	14944406
sigcomm2004-98.93.251-20040831-1400	609592181
sigcomm2004-98.93.251-20040831-1900	56282052
sigcomm2004-98.93.251-20040831-2000	52018520
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0
sigcomm2004-98.93.251-20040902-1400	807477624
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	8870
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	784
sotm27-172.16.134-20030302-1800	380
sotm27-172.16.134-20030302-1900	0
sotm27-172.16.134-20030304-1400	11889794
sotm27-172.16.134-20030305-0000	11895691
sotm27-172.16.134-20030305-1000	184838
ten-64.5.53.67-20041216-2117	24726
ten-64.5.53.67-20050107-2326	51812
ten-64.5.53.67-20070804-1200	1624275
ten-64.5.53.67-20070806-0000	20924179
ten-64.5.53.67-20070806-0900	18191
ten-64.5.53.67-20070806-1000	89888
ten-64.5.53.67-20070806-1800	1330346
ten-64.5.53.67-20070806-1900	253988
ten-64.5.53.67-20070813-0000	18833674
ten-64.5.53.67-20070813-1000	169324
ten-64.5.53.67-20070813-1800	229891
ten-64.5.53.67-20071105-0000	2726257
ten-64.5.53.67-20071105-1800	117095
ten-64.5.53.67-20080302-0000	3971929
ten-64.5.53.67-20080302-1800	155431

Table K.6: Bytes out counts for all traces.

Trace	Bytes out
ames-173.148.249-20020814-1600	40977
ames-237.24.63-20020814-1600	0
ames-237.24.63-20020814-1700	0
ames-237.24.65-20020814-1600	0
ames-239.65.134-20020814-1600	0
ames-239.65.134-20020814-1700	0
ames-3.12.2-20020814-1600	0
ames-3.3.31-20020814-1600	59387382
dartmouth-190.84.116-20031103-0000	1969090263
dartmouth-190.84.116-20031103-1800	2339434
dartmouth-190.84.172-20031103-0000	193455502
dartmouth-190.84.172-20031103-0900	213694
dartmouth-190.84.172-20031103-1000	185345
dartmouth-190.84.172-20031103-1800	4537378
dartmouth-190.84.172-20031110-0000	4233578626
dartmouth-190.84.172-20031110-1000	6437309
dartmouth-190.84.172-20031110-1800	203869410
dartmouth-190.84.172-20040127-1400	134060654
dartmouth-190.84.172-20040128-0000	46555088
dartmouth-190.84.172-20040129-1800	20469720
dartmouth-190.84.224-20031103-0000	40624344
dartmouth-190.84.44-20031103-0000	579183996
dartmouth-190.84.44-20031103-1800	26261992
dartmouth-190.84.69-20031103-0000	2586787358
dartmouth-190.84.69-20031103-0900	35577
dartmouth-190.84.69-20031103-1000	41891
dartmouth-190.84.69-20031103-1800	195321518
dartmouth-190.84.69-20031110-0000	2669464002
dartmouth-190.84.69-20031110-1000	43260143
Continued on next page	

Table K.6 – continued from previous page

Trace	Bytes out
dartmouth-190.84.69-20031110-1800	116294420
dc10-192.168.2-20020803-1200	4670676
dsl1-69.110.74.139-20070201-0000	3564533
dsl1-69.110.74.139-20070201-1800	34530
dsl1-69.110.79.251-20070806-0000	4799992
dsl1-69.110.79.251-20070806-0900	61832
dsl1-69.110.79.251-20070806-1000	339625
dsl1-69.110.79.251-20070806-1800	142534
dsl1-69.110.79.251-20070806-1900	59138
dsl1-69.110.79.251-20070813-0000	16191389
dsl1-69.110.79.251-20070813-1000	669375
dsl1-69.110.79.251-20070813-1800	80365
dsl1-69.225.88.159-20070302-0000	25333961
dsl1-69.225.88.159-20070302-1800	14284
dsl1-71.133.175.192-20070804-1200	1419841
lbl-128.3.23-20041216-2117	249363372
lbl-128.3.23-20050107-2326	97874610
lbl-128.3.45-20041216-2117	3127580
lbl-128.3.46-20041216-2117	2483758
lbl-128.3.47-20041216-2117	2017214
lbl-131.243.12-20041215-1812	31281403
lbl-131.243.13-20041215-1812	62913250
lbl-131.243.140-20041215-1109	12016797
lbl-131.243.140-20041215-1942	1252993902
lbl-131.243.155-20041216-2117	40819728
lbl-131.243.155-20050106-2225	4742869
sigcomm2004-98.93.251-20040831-1400	117062069
sigcomm2004-98.93.251-20040831-1900	7611390
sigcomm2004-98.93.251-20040831-2000	21084430
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0

Continued on next page

Table K.6 – continued from previous page

Trace	Bytes out
sigcomm2004-98.93.251-20040902-1400	113233500
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	7655
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	443
sotm27-172.16.134-20030302-1800	443
sotm27-172.16.134-20030302-1900	0
sotm27-172.16.134-20030304-1400	842096
sotm27-172.16.134-20030305-0000	840453
sotm27-172.16.134-20030305-1000	140234
ten-64.5.53.67-20041216-2117	5385
ten-64.5.53.67-20050107-2326	37007
ten-64.5.53.67-20070804-1200	511416
ten-64.5.53.67-20070806-0000	11525892
ten-64.5.53.67-20070806-0900	6505
ten-64.5.53.67-20070806-1000	36287
ten-64.5.53.67-20070806-1800	9213683
ten-64.5.53.67-20070806-1900	105516
ten-64.5.53.67-20070813-0000	1363131
ten-64.5.53.67-20070813-1000	55605
ten-64.5.53.67-20070813-1800	51685
ten-64.5.53.67-20071105-0000	3987141
ten-64.5.53.67-20071105-1800	53723
ten-64.5.53.67-20080302-0000	4468479
ten-64.5.53.67-20080302-1800	85896

Table K.7: SYN-ONLY rate ratios for all traces.

Trace	SYN-ONLY rate
ames-173.148.249-20020814-1600	1
ames-237.24.63-20020814-1600	1
ames-237.24.63-20020814-1700	0
ames-237.24.65-20020814-1600	1
ames-239.65.134-20020814-1600	1
ames-239.65.134-20020814-1700	1
ames-3.12.2-20020814-1600	1
ames-3.3.31-20020814-1600	1
dartmouth-190.84.116-20031103-0000	0.138383898004958
dartmouth-190.84.116-20031103-1800	0.158383035122598
dartmouth-190.84.172-20031103-0000	0.146033077660594
dartmouth-190.84.172-20031103-0900	0.00906344410876133
dartmouth-190.84.172-20031103-1000	0.0244648318042813
dartmouth-190.84.172-20031103-1800	0.0949788263762855
dartmouth-190.84.172-20031110-0000	0.103907766669748
dartmouth-190.84.172-20031110-1000	0.0483333333333333
dartmouth-190.84.172-20031110-1800	0.121554770318021
dartmouth-190.84.172-20040127-1400	0.00860305519812597
dartmouth-190.84.172-20040128-0000	0.0421077119925555
dartmouth-190.84.172-20040129-1800	0.0384316770186335
dartmouth-190.84.224-20031103-0000	0.296218863867153
dartmouth-190.84.44-20031103-0000	0.0852628039896896
dartmouth-190.84.44-20031103-1800	0.0478629111680126
dartmouth-190.84.69-20031103-0000	0.932584931706715
dartmouth-190.84.69-20031103-0900	0
dartmouth-190.84.69-20031103-1000	0
dartmouth-190.84.69-20031103-1800	0.937887590573797
dartmouth-190.84.69-20031110-0000	0.889623399605357
dartmouth-190.84.69-20031110-1000	0.994685200327065
dartmouth-190.84.69-20031110-1800	0.843160229317488
Continued on next page	

Table K.7 – continued from previous page

Trace	SYN-ONLY rate
dc10-192.168.2-20020803-1200	0.0979102642901045
dsl1-69.110.74.139-20070201-0000	0.0633626097867001
dsl1-69.110.74.139-20070201-1800	0.297297297297297
dsl1-69.110.79.251-20070806-0000	0.11451048951049
dsl1-69.110.79.251-20070806-0900	0.0697674418604651
dsl1-69.110.79.251-20070806-1000	0.0421052631578947
dsl1-69.110.79.251-20070806-1800	0.3888888888888889
dsl1-69.110.79.251-20070806-1900	0.47887323943662
dsl1-69.110.79.251-20070813-0000	0.0282174810736407
dsl1-69.110.79.251-20070813-1000	0.0915032679738562
dsl1-69.110.79.251-20070813-1800	0.215686274509804
dsl1-69.225.88.159-20070302-0000	0.0561035758323058
dsl1-69.225.88.159-20070302-1800	1
dsl1-71.133.175.192-20070804-1200	0.053347280334728
lbl-128.3.23-20041216-2117	0.00317994762439207
lbl-128.3.23-20050107-2326	0.0326295585412668
lbl-128.3.45-20041216-2117	0.0275229357798165
lbl-128.3.46-20041216-2117	0.0632603406326034
lbl-128.3.47-20041216-2117	0.053921568627451
lbl-131.243.12-20041215-1812	0.0092378752886836
lbl-131.243.13-20041215-1812	0.0204657727593507
lbl-131.243.140-20041215-1109	0.00375234521575985
lbl-131.243.140-20041215-1942	0.00760382140768181
lbl-131.243.155-20041216-2117	0.00345224395857307
lbl-131.243.155-20050106-2225	0.00181653042688465
sigcomm2004-98.93.251-20040831-1400	0.261985988425221
sigcomm2004-98.93.251-20040831-1900	0.0453001132502831
sigcomm2004-98.93.251-20040831-2000	0.139587073608618
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0
sigcomm2004-98.93.251-20040902-1400	0.103674440012274
Continued on next page	

Table K.7 – continued from previous page

Trace	SYN-ONLY rate
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	0
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	0
sotm27-172.16.134-20030302-1800	0
sotm27-172.16.134-20030302-1900	0
sotm27-172.16.134-20030304-1400	0.0105633802816901
sotm27-172.16.134-20030305-0000	0.0105932203389831
sotm27-172.16.134-20030305-1000	0.0160349854227405
ten-64.5.53.67-20041216-2117	0.6666666666666667
ten-64.5.53.67-20050107-2326	0.5
ten-64.5.53.67-20070804-1200	0.215189873417722
ten-64.5.53.67-20070806-0000	0.17607223476298
ten-64.5.53.67-20070806-0900	0.272727272727273
ten-64.5.53.67-20070806-1000	0.117647058823529
ten-64.5.53.67-20070806-1800	0.3333333333333333
ten-64.5.53.67-20070806-1900	0.125
ten-64.5.53.67-20070813-0000	0.161290322580645
ten-64.5.53.67-20070813-1000	0.0625
ten-64.5.53.67-20070813-1800	0.0454545454545455
ten-64.5.53.67-20071105-0000	0.22673434856176
ten-64.5.53.67-20071105-1800	0.13636363636363636
ten-64.5.53.67-20080302-0000	0.0913294797687861
ten-64.5.53.67-20080302-1800	0.125

Table K.8: SYN-ACK rate ratios for all traces.

Trace	SYN-ACK rate
ames-173.148.249-20020814-1600	0
Continued on next page	

Table K.8 – continued from previous page

Trace	SYN-ACK rate
ames-237.24.63-20020814-1600	0
ames-237.24.63-20020814-1700	0
ames-237.24.65-20020814-1600	0
ames-239.65.134-20020814-1600	0
ames-239.65.134-20020814-1700	0
ames-3.12.2-20020814-1600	0
ames-3.3.31-20020814-1600	0
dartmouth-190.84.116-20031103-0000	0.000531224176602526
dartmouth-190.84.116-20031103-1800	0.00132538104705103
dartmouth-190.84.172-20031103-0000	9.98721636305529e-05
dartmouth-190.84.172-20031103-0900	0
dartmouth-190.84.172-20031103-1000	0
dartmouth-190.84.172-20031103-1800	0.000302480338777979
dartmouth-190.84.172-20031110-0000	0.000330159400958783
dartmouth-190.84.172-20031110-1000	0
dartmouth-190.84.172-20031110-1800	0
dartmouth-190.84.172-20040127-1400	4.12418753505559e-05
dartmouth-190.84.172-20040128-0000	0.0004652785855531
dartmouth-190.84.172-20040129-1800	0
dartmouth-190.84.224-20031103-0000	0.000477868705573144
dartmouth-190.84.44-20031103-0000	0.000526728678695506
dartmouth-190.84.44-20031103-1800	0
dartmouth-190.84.69-20031103-0000	1.12232185946286e-05
dartmouth-190.84.69-20031103-0900	0
dartmouth-190.84.69-20031103-1000	0
dartmouth-190.84.69-20031103-1800	0
dartmouth-190.84.69-20031110-0000	2.81578485670904e-05
dartmouth-190.84.69-20031110-1000	0
dartmouth-190.84.69-20031110-1800	4.18462568523246e-05
dc10-192.168.2-20020803-1200	0.000122925629993854
dsl1-69.110.74.139-20070201-0000	0.00439146800501882
Continued on next page	

Table K.8 – continued from previous page

Trace	SYN-ACK rate
dsl1-69.110.74.139-20070201-1800	0
dsl1-69.110.79.251-20070806-0000	0
dsl1-69.110.79.251-20070806-0900	0
dsl1-69.110.79.251-20070806-1000	0
dsl1-69.110.79.251-20070806-1800	0
dsl1-69.110.79.251-20070806-1900	0
dsl1-69.110.79.251-20070813-0000	0
dsl1-69.110.79.251-20070813-1000	0
dsl1-69.110.79.251-20070813-1800	0
dsl1-69.225.88.159-20070302-0000	0.00739827373612824
dsl1-69.225.88.159-20070302-1800	0
dsl1-71.133.175.192-20070804-1200	0
lbl-128.3.23-20041216-2117	0.000561167227833895
lbl-128.3.23-20050107-2326	0.00127959053103007
lbl-128.3.45-20041216-2117	0
lbl-128.3.46-20041216-2117	0.0024330900243309
lbl-128.3.47-20041216-2117	0
lbl-131.243.12-20041215-1812	0.000769822940723634
lbl-131.243.13-20041215-1812	0
lbl-131.243.140-20041215-1109	0.024390243902439
lbl-131.243.140-20041215-1942	0.00487424449210372
lbl-131.243.155-20041216-2117	0
lbl-131.243.155-20050106-2225	0
sigcomm2004-98.93.251-20040831-1400	0.0139201949436491
sigcomm2004-98.93.251-20040831-1900	0.00339750849377123
sigcomm2004-98.93.251-20040831-2000	0.000448833034111311
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0
sigcomm2004-98.93.251-20040902-1400	0.00717244553544032
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	0
Continued on next page	

Table K.8 – continued from previous page

Trace	SYN-ACK rate
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	0
sotm27-172.16.134-20030302-1800	0
sotm27-172.16.134-20030302-1900	0
sotm27-172.16.134-20030304-1400	0.0119718309859155
sotm27-172.16.134-20030305-0000	0.0120056497175141
sotm27-172.16.134-20030305-1000	0.021865889212828
ten-64.5.53.67-20041216-2117	0
ten-64.5.53.67-20050107-2326	0
ten-64.5.53.67-20070804-1200	0
ten-64.5.53.67-20070806-0000	0.00225733634311512
ten-64.5.53.67-20070806-0900	0
ten-64.5.53.67-20070806-1000	0
ten-64.5.53.67-20070806-1800	0
ten-64.5.53.67-20070806-1900	0
ten-64.5.53.67-20070813-0000	0.00569259962049336
ten-64.5.53.67-20070813-1000	0
ten-64.5.53.67-20070813-1800	0
ten-64.5.53.67-20071105-0000	0
ten-64.5.53.67-20071105-1800	0
ten-64.5.53.67-20080302-0000	0.00578034682080925
ten-64.5.53.67-20080302-1800	0

Table K.9: Idle connection rate ratios for all traces.

Trace	Idle connection rate
ames-173.148.249-20020814-1600	0
ames-237.24.63-20020814-1600	0
ames-237.24.63-20020814-1700	0
Continued on next page	

Table K.9 – continued from previous page

Trace	Idle connection rate
ames-237.24.65-20020814-1600	0
ames-239.65.134-20020814-1600	0
ames-239.65.134-20020814-1700	0
ames-3.12.2-20020814-1600	0
ames-3.3.31-20020814-1600	0
dartmouth-190.84.116-20031103-0000	0.00247904615747846
dartmouth-190.84.116-20031103-1800	0.00596421471172962
dartmouth-190.84.172-20031103-0000	0.000818951741770534
dartmouth-190.84.172-20031103-0900	0.00302114803625378
dartmouth-190.84.172-20031103-1000	0.00305810397553517
dartmouth-190.84.172-20031103-1800	0.000604960677555959
dartmouth-190.84.172-20031110-0000	0.00438451684473264
dartmouth-190.84.172-20031110-1000	0.01333333333333333
dartmouth-190.84.172-20031110-1800	0.00424028268551237
dartmouth-190.84.172-20040127-1400	0.000255699627173447
dartmouth-190.84.172-20040128-0000	0.000814237524717925
dartmouth-190.84.172-20040129-1800	0.00892857142857143
dartmouth-190.84.224-20031103-0000	0.00310614658622543
dartmouth-190.84.44-20031103-0000	0.00494228398520677
dartmouth-190.84.44-20031103-1800	0.00374236753988576
dartmouth-190.84.69-20031103-0000	0.000364754604325428
dartmouth-190.84.69-20031103-0900	0
dartmouth-190.84.69-20031103-1000	0
dartmouth-190.84.69-20031103-1800	0.000342711665252301
dartmouth-190.84.69-20031110-0000	0.000589148831557584
dartmouth-190.84.69-20031110-1000	0
dartmouth-190.84.69-20031110-1800	0.00175754278779763
dc10-192.168.2-20020803-1200	0.00522433927473878
dsl1-69.110.74.139-20070201-0000	0
dsl1-69.110.74.139-20070201-1800	0
dsl1-69.110.79.251-20070806-0000	0.000874125874125874
Continued on next page	

Table K.9 – continued from previous page

Trace	Idle connection rate
dsl1-69.110.79.251-20070806-0900	0
dsl1-69.110.79.251-20070806-1000	0
dsl1-69.110.79.251-20070806-1800	0
dsl1-69.110.79.251-20070806-1900	0
dsl1-69.110.79.251-20070813-0000	0.000458820830465703
dsl1-69.110.79.251-20070813-1000	0
dsl1-69.110.79.251-20070813-1800	0
dsl1-69.225.88.159-20070302-0000	0
dsl1-69.225.88.159-20070302-1800	0
dsl1-71.133.175.192-20070804-1200	0
lbl-128.3.23-20041216-2117	0.00224466891133558
lbl-128.3.23-20050107-2326	0.00319897632757518
lbl-128.3.45-20041216-2117	0
lbl-128.3.46-20041216-2117	0.0048661800486618
lbl-128.3.47-20041216-2117	0
lbl-131.243.12-20041215-1812	0.000769822940723634
lbl-131.243.13-20041215-1812	0.00282286520818631
lbl-131.243.140-20041215-1109	0
lbl-131.243.140-20041215-1942	0.00799376096705011
lbl-131.243.155-20041216-2117	0.0126582278481013
lbl-131.243.155-20050106-2225	0.00272479564032698
sigcomm2004-98.93.251-20040831-1400	0.00727992689613159
sigcomm2004-98.93.251-20040831-1900	0.0192525481313703
sigcomm2004-98.93.251-20040831-2000	0.00583482944344704
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0
sigcomm2004-98.93.251-20040902-1400	0.0111230438784903
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	0
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	0
Continued on next page	

Table K.9 – continued from previous page

Trace	Idle connection rate
sotm27-172.16.134-20030302-1800	0
sotm27-172.16.134-20030302-1900	0
sotm27-172.16.134-20030304-1400	0.0359154929577465
sotm27-172.16.134-20030305-0000	0.0367231638418079
sotm27-172.16.134-20030305-1000	0.0816326530612245
ten-64.5.53.67-20041216-2117	0
ten-64.5.53.67-20050107-2326	0
ten-64.5.53.67-20070804-1200	0.00316455696202532
ten-64.5.53.67-20070806-0000	0
ten-64.5.53.67-20070806-0900	0
ten-64.5.53.67-20070806-1000	0
ten-64.5.53.67-20070806-1800	0
ten-64.5.53.67-20070806-1900	0
ten-64.5.53.67-20070813-0000	0
ten-64.5.53.67-20070813-1000	0
ten-64.5.53.67-20070813-1800	0
ten-64.5.53.67-20071105-0000	0.00676818950930626
ten-64.5.53.67-20071105-1800	0.0454545454545455
ten-64.5.53.67-20080302-0000	0.00115606936416185
ten-64.5.53.67-20080302-1800	0

Table K.10: Half-open connection rate ratios for all traces.

Trace	Half-open connection rate
ames-173.148.249-20020814-1600	1
ames-237.24.63-20020814-1600	1
ames-237.24.63-20020814-1700	0
ames-237.24.65-20020814-1600	1
ames-239.65.134-20020814-1600	1
Continued on next page	

Table K.10 – continued from previous page

Trace	Half-open connection rate
ames-239.65.134-20020814-1700	1
ames-3.12.2-20020814-1600	1
ames-3.3.31-20020814-1600	1
dartmouth-190.84.116-20031103-0000	0.23409278715618
dartmouth-190.84.116-20031103-1800	0.275679257786614
dartmouth-190.84.172-20031103-0000	0.18174736337488
dartmouth-190.84.172-20031103-0900	0.0181268882175227
dartmouth-190.84.172-20031103-1000	0.0336391437308868
dartmouth-190.84.172-20031103-1800	0.139745916515427
dartmouth-190.84.172-20031110-0000	0.128181085828238
dartmouth-190.84.172-20031110-1000	0.0633333333333333
dartmouth-190.84.172-20031110-1800	0.170671378091873
dartmouth-190.84.172-20040127-1400	0.0227572668184368
dartmouth-190.84.172-20040128-0000	0.232871932069327
dartmouth-190.84.172-20040129-1800	0.355201863354037
dartmouth-190.84.224-20031103-0000	0.359894868884774
dartmouth-190.84.44-20031103-0000	0.164922111397512
dartmouth-190.84.44-20031103-1800	0.141028166239905
dartmouth-190.84.69-20031103-0000	0.94702219952638
dartmouth-190.84.69-20031103-0900	0
dartmouth-190.84.69-20031103-1000	0
dartmouth-190.84.69-20031103-1800	0.946177948952281
dartmouth-190.84.69-20031110-0000	0.912294799678567
dartmouth-190.84.69-20031110-1000	0.994685200327065
dartmouth-190.84.69-20031110-1800	0.873331380508014
dc10-192.168.2-20020803-1200	0.146650276582667
dsl1-69.110.74.139-20070201-0000	0.269134253450439
dsl1-69.110.74.139-20070201-1800	0.513513513513513
dsl1-69.110.79.251-20070806-0000	0.287587412587413
dsl1-69.110.79.251-20070806-0900	0.27906976744186
dsl1-69.110.79.251-20070806-1000	0.221052631578947
Continued on next page	

Table K.10 – continued from previous page

Trace	Half-open connection rate
dsl1-69.110.79.251-20070806-1800	0.4555555555555556
dsl1-69.110.79.251-20070806-1900	0.591549295774648
dsl1-69.110.79.251-20070813-0000	0.123422803395274
dsl1-69.110.79.251-20070813-1000	0.254901960784314
dsl1-69.110.79.251-20070813-1800	0.372549019607843
dsl1-69.225.88.159-20070302-0000	0.302712700369914
dsl1-69.225.88.159-20070302-1800	1
dsl1-71.133.175.192-20070804-1200	0.27092050209205
lbl-128.3.23-20041216-2117	0.335203890759446
lbl-128.3.23-20050107-2326	0.167306461932182
lbl-128.3.45-20041216-2117	0.288990825688073
lbl-128.3.46-20041216-2117	0.250608272506083
lbl-128.3.47-20041216-2117	0.0637254901960784
lbl-131.243.12-20041215-1812	0.141647421093149
lbl-131.243.13-20041215-1812	0.242766407904023
lbl-131.243.140-20041215-1109	0.0365853658536585
lbl-131.243.140-20041215-1942	0.102944043673231
lbl-131.243.155-20041216-2117	0.108170310701956
lbl-131.243.155-20050106-2225	0.157129881925522
sigcomm2004-98.93.251-20040831-1400	0.401644837039293
sigcomm2004-98.93.251-20040831-1900	0.300113250283126
sigcomm2004-98.93.251-20040831-2000	0.281418312387792
sigcomm2004-98.93.251-20040901-0900	0
sigcomm2004-98.93.251-20040901-1000	0
sigcomm2004-98.93.251-20040902-1400	0.269062595888309
sigcomm2004-98.93.251-20040903-1000	0
sotm27-172.16.134-20030302-0000	0
sotm27-172.16.134-20030302-0900	0
sotm27-172.16.134-20030302-1000	0
sotm27-172.16.134-20030302-1800	0
sotm27-172.16.134-20030302-1900	0

Continued on next page

Table K.10 – continued from previous page

Trace	Half-open connection rate
sotm27-172.16.134-20030304-1400	0.457042253521127
sotm27-172.16.134-20030305-0000	0.459039548022599
sotm27-172.16.134-20030305-1000	0.137026239067055
ten-64.5.53.67-20041216-2117	0.6666666666666667
ten-64.5.53.67-20050107-2326	0.5
ten-64.5.53.67-20070804-1200	0.234177215189873
ten-64.5.53.67-20070806-0000	0.182844243792325
ten-64.5.53.67-20070806-0900	0.272727272727273
ten-64.5.53.67-20070806-1000	0.117647058823529
ten-64.5.53.67-20070806-1800	0.333333333333333
ten-64.5.53.67-20070806-1900	0.125
ten-64.5.53.67-20070813-0000	0.176470588235294
ten-64.5.53.67-20070813-1000	0.0625
ten-64.5.53.67-20070813-1800	0.0909090909090909
ten-64.5.53.67-20071105-0000	0.238578680203046
ten-64.5.53.67-20071105-1800	0.181818181818182
ten-64.5.53.67-20080302-0000	0.10635838150289
ten-64.5.53.67-20080302-1800	0.125

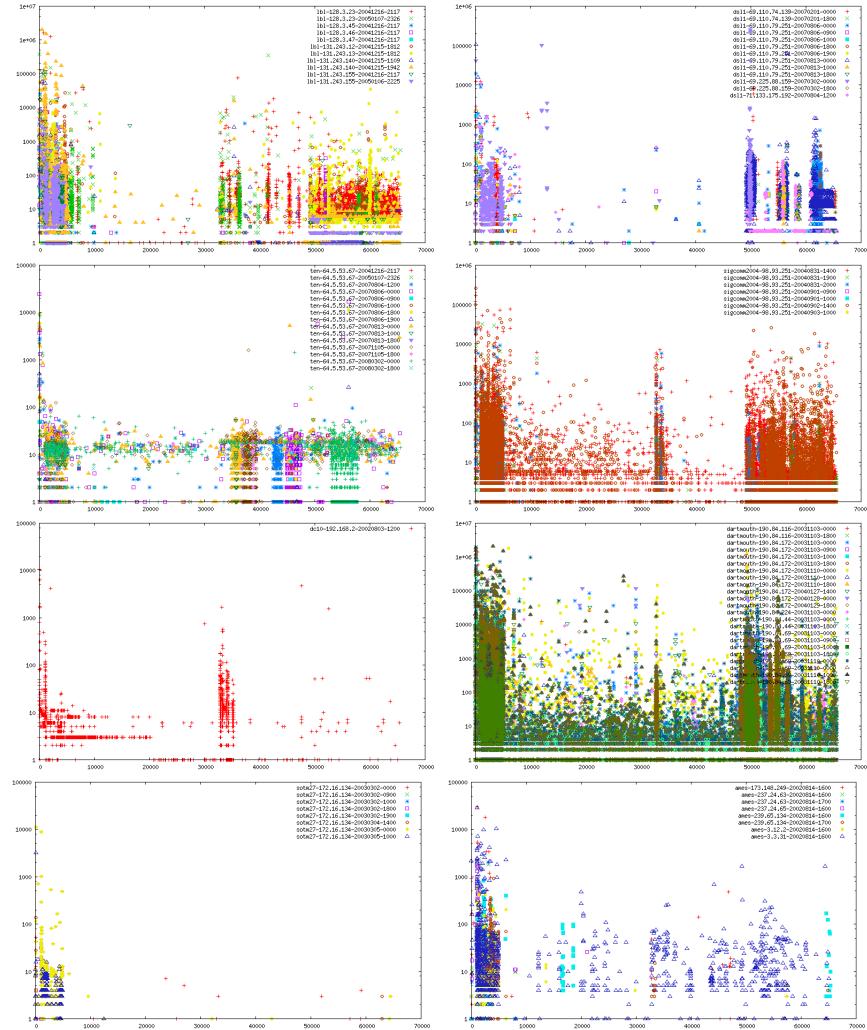


Figure K.1: Packet Service distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

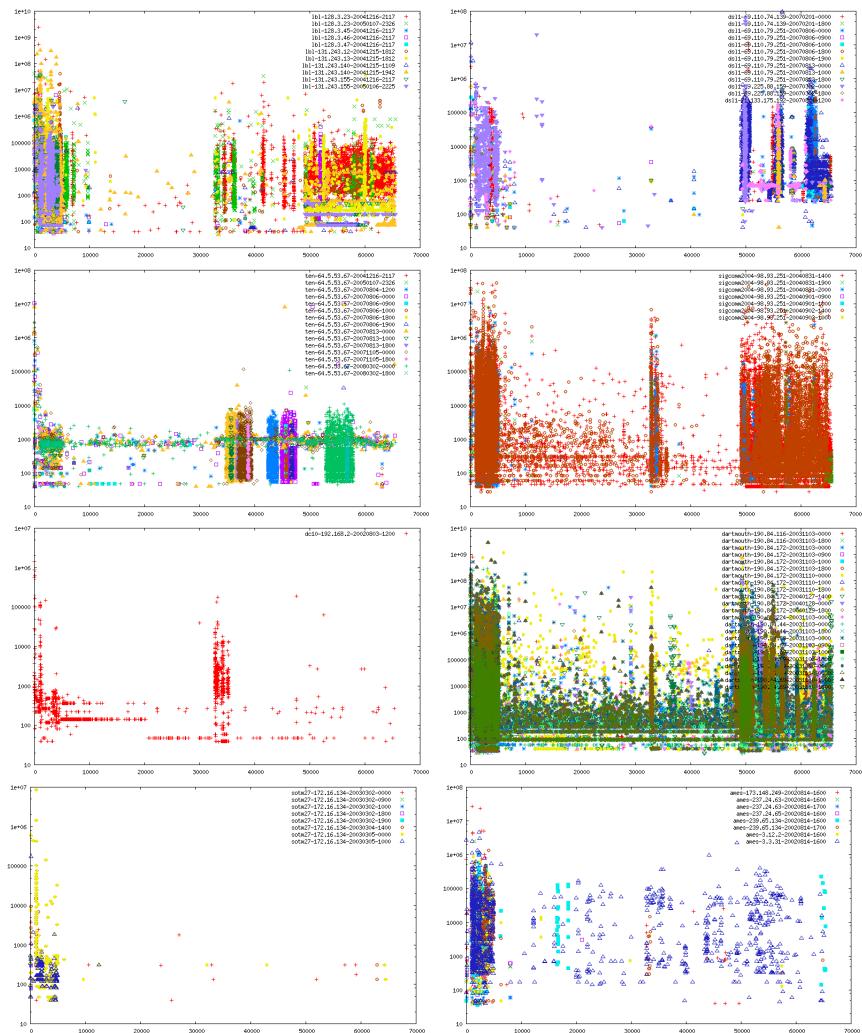


Figure K.2: Bytes Service distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

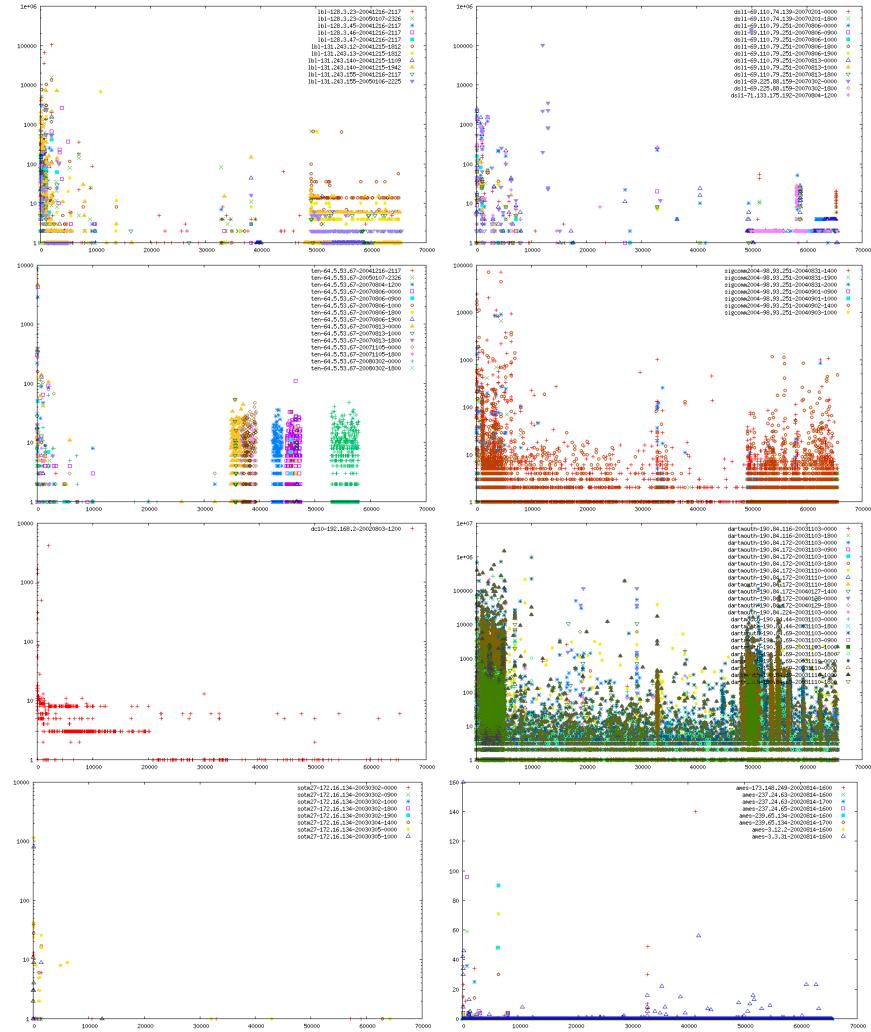


Figure K.3: Connection Service distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

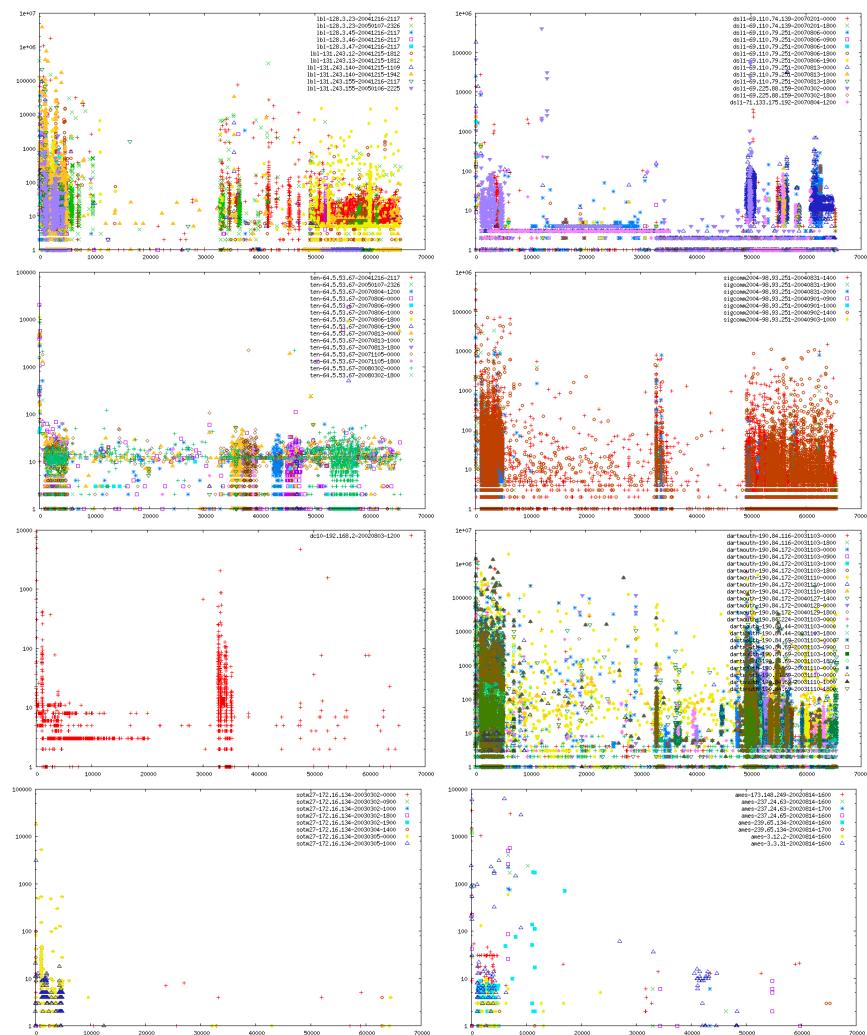


Figure K.4: Packet Source port distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

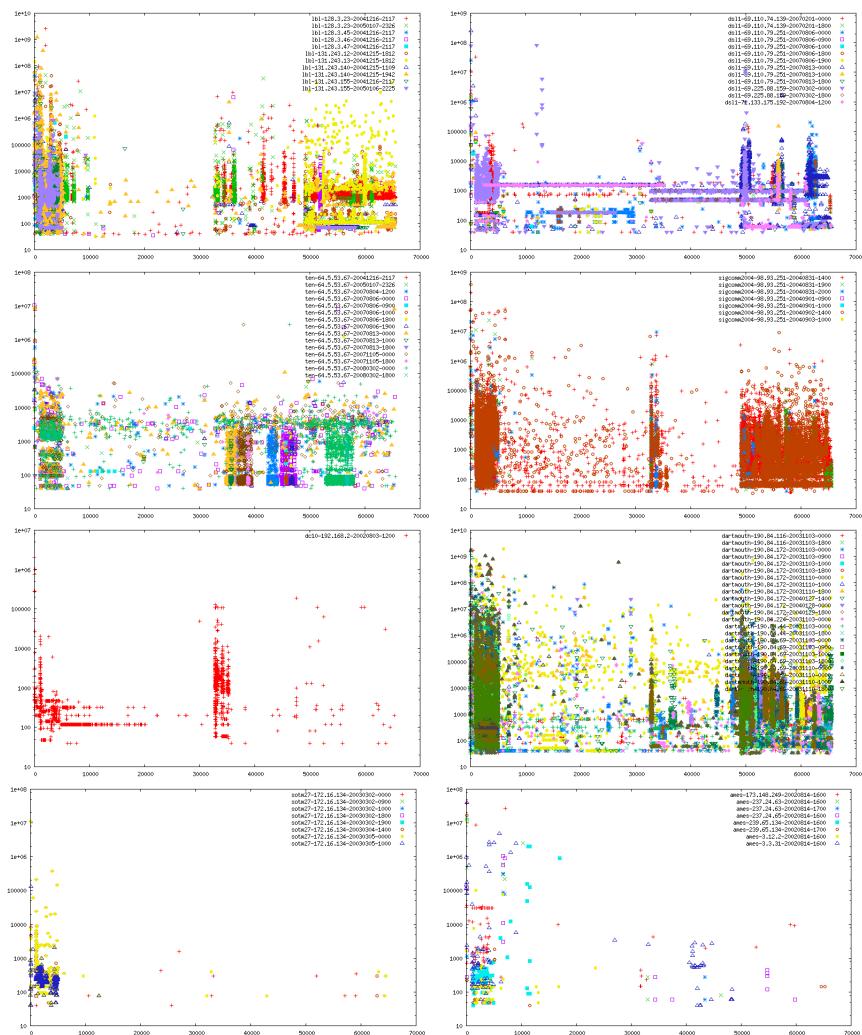


Figure K.5: Bytes Source port distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

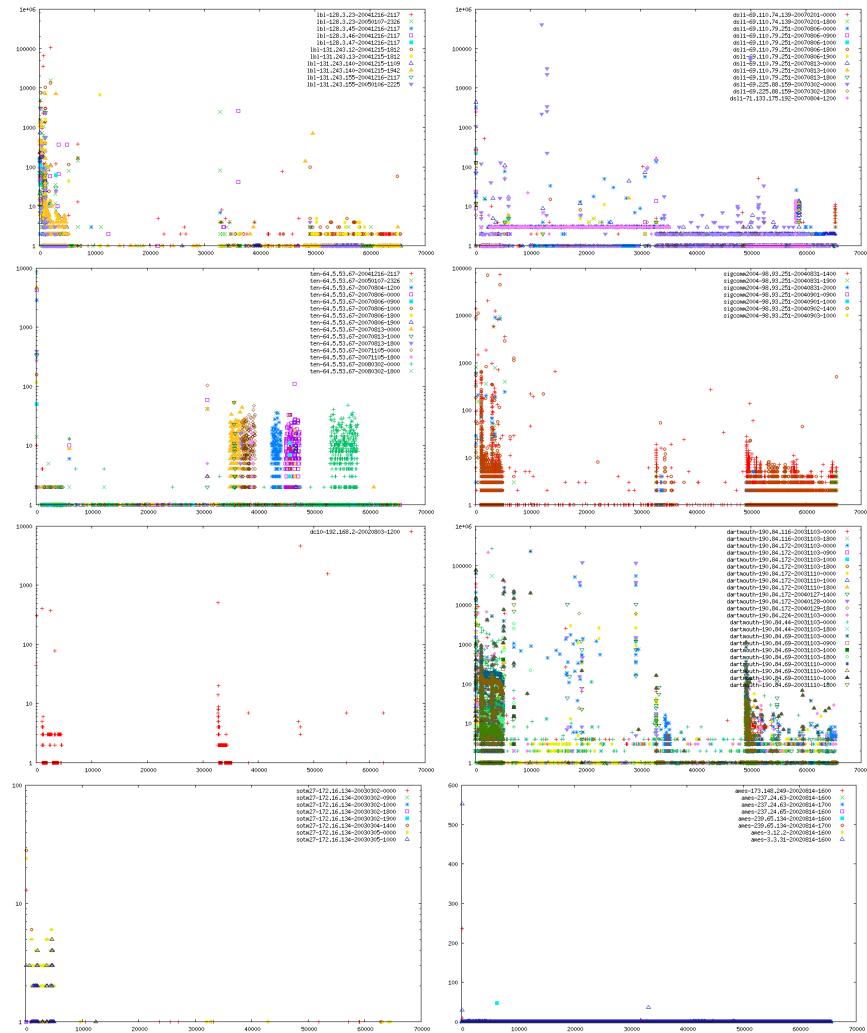


Figure K.6: Connection Source port distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

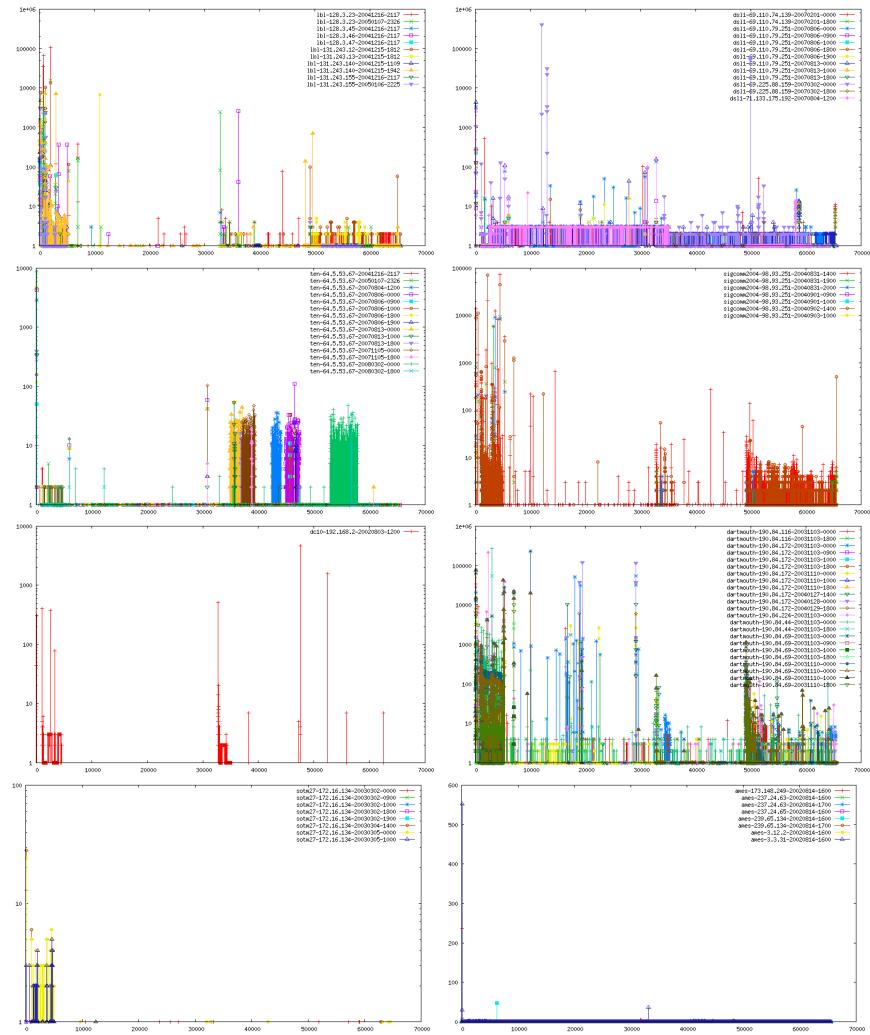


Figure K.7: Connection Source port distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

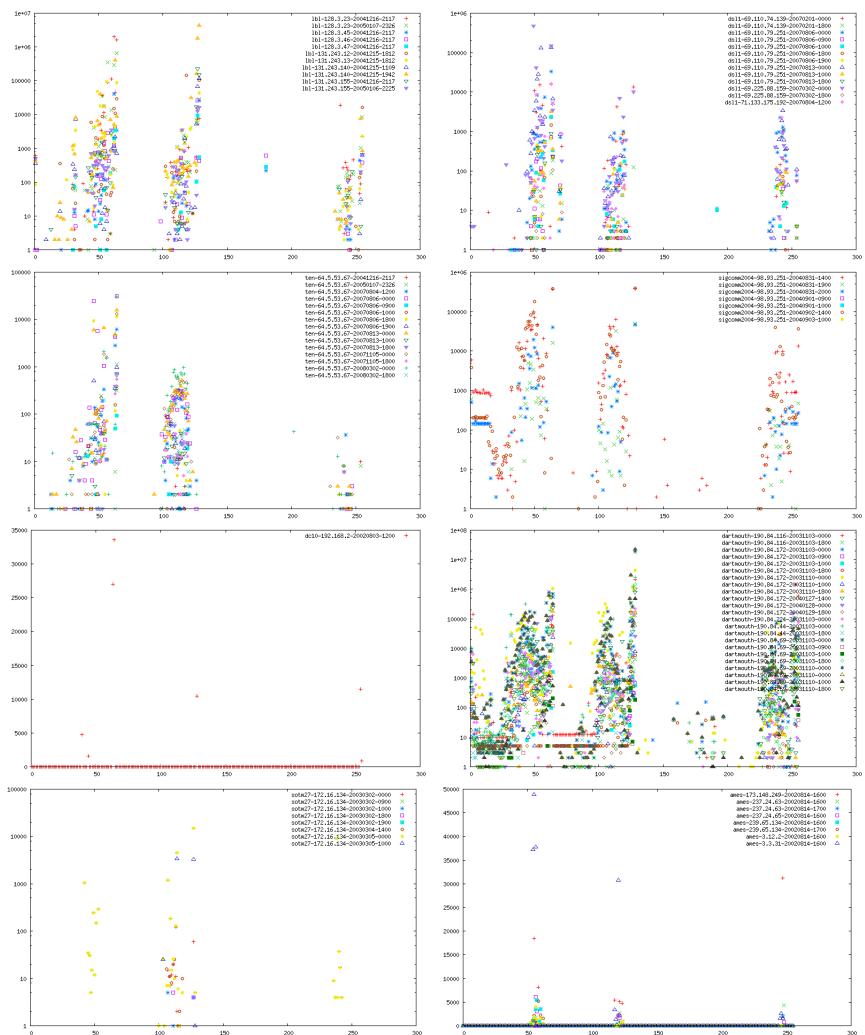


Figure K.8: Packet TTL distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

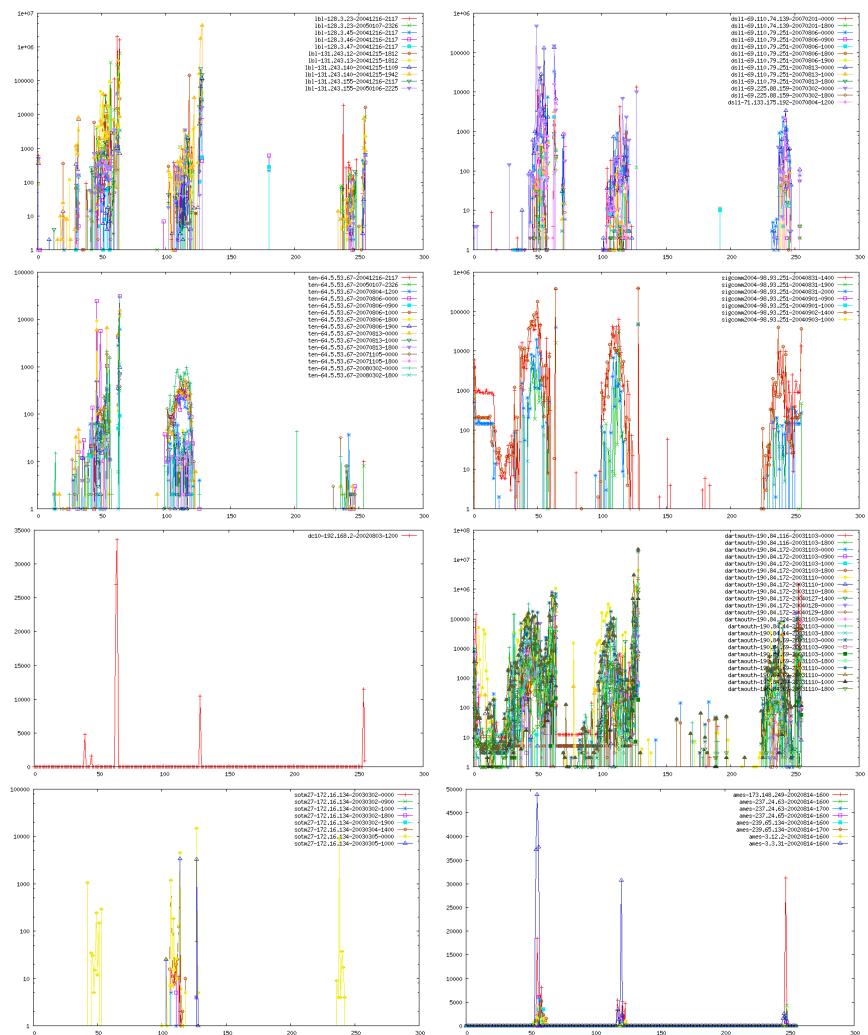


Figure K.9: Packet TTL distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

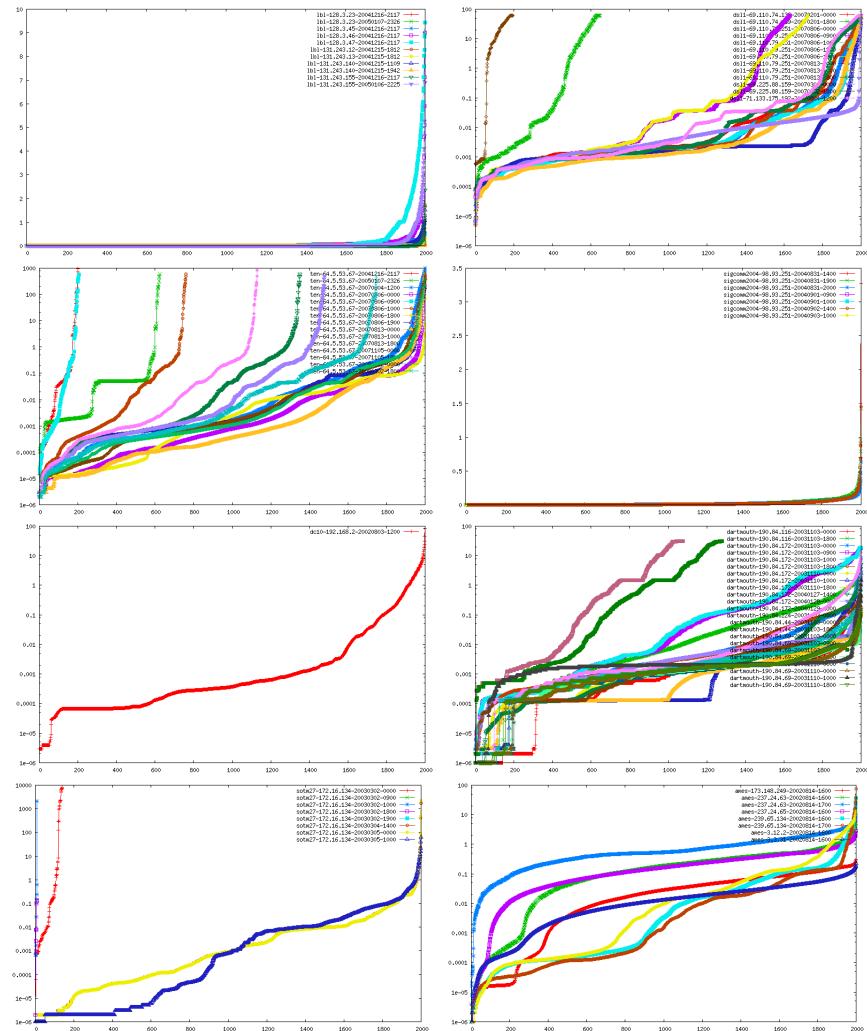


Figure K.10: InterPacket delta distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

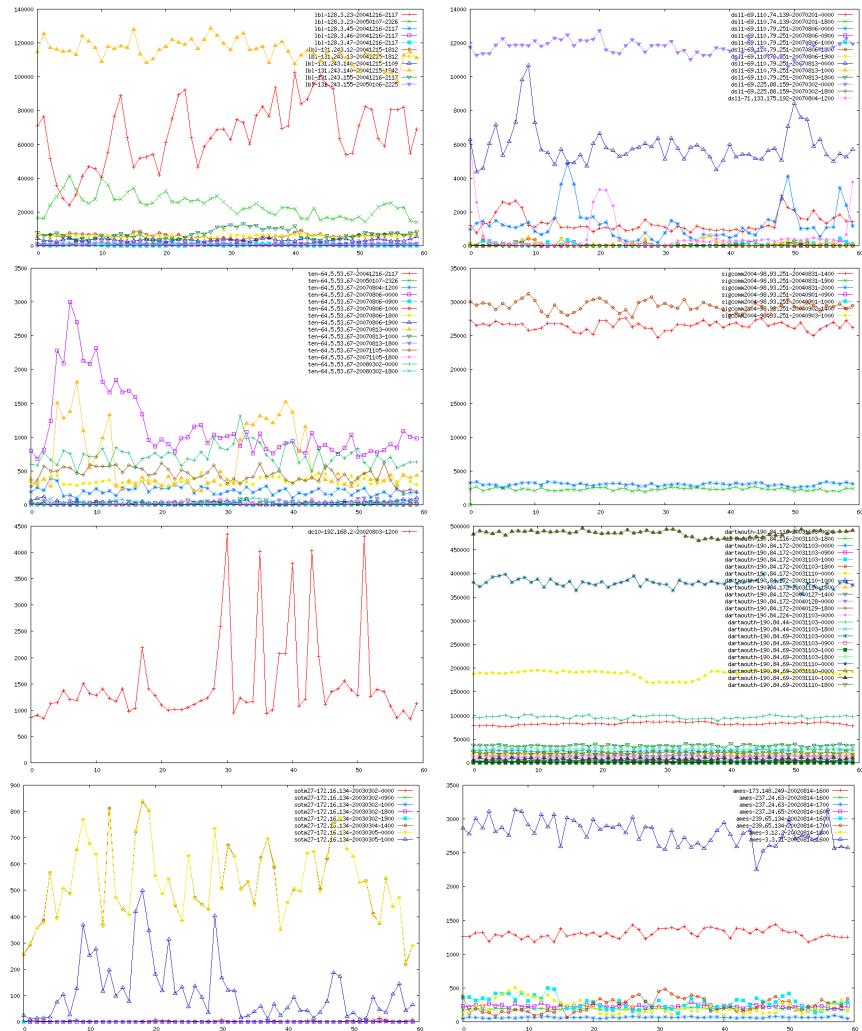


Figure K.11: Packet sec distributions for all traces, with all the traces from the same dataset plotted together.



Figure K.12: Packet min distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

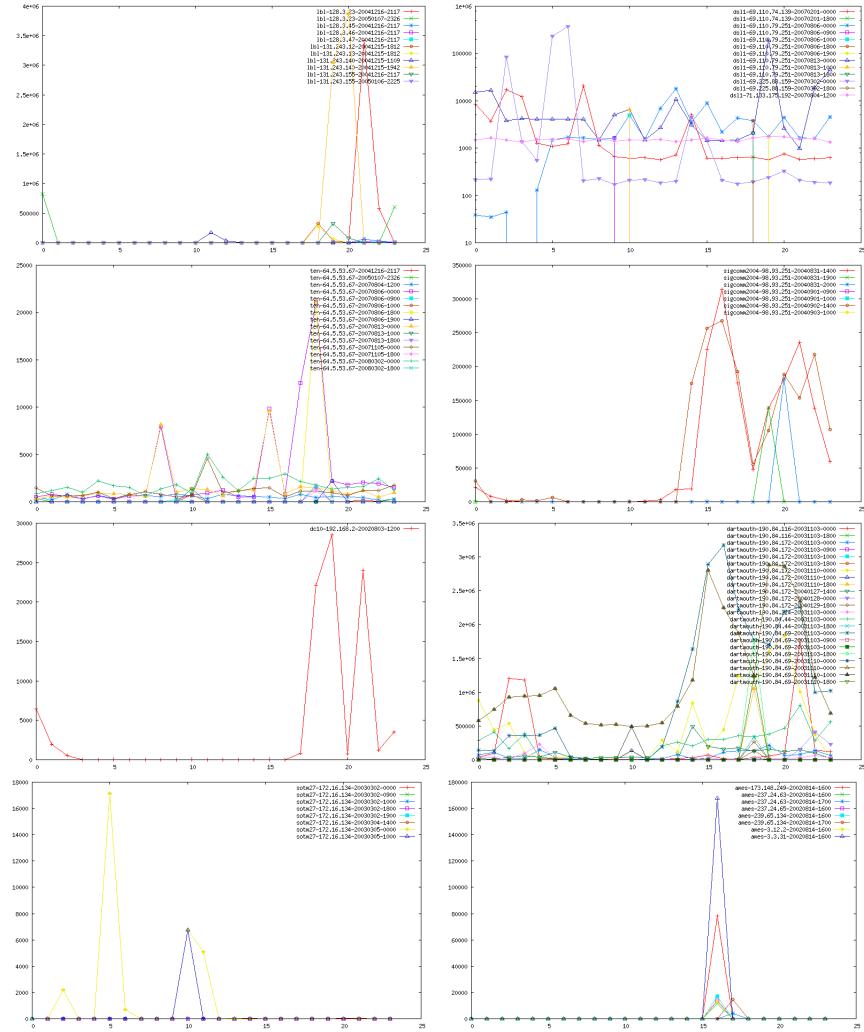


Figure K.13: Packet GmHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

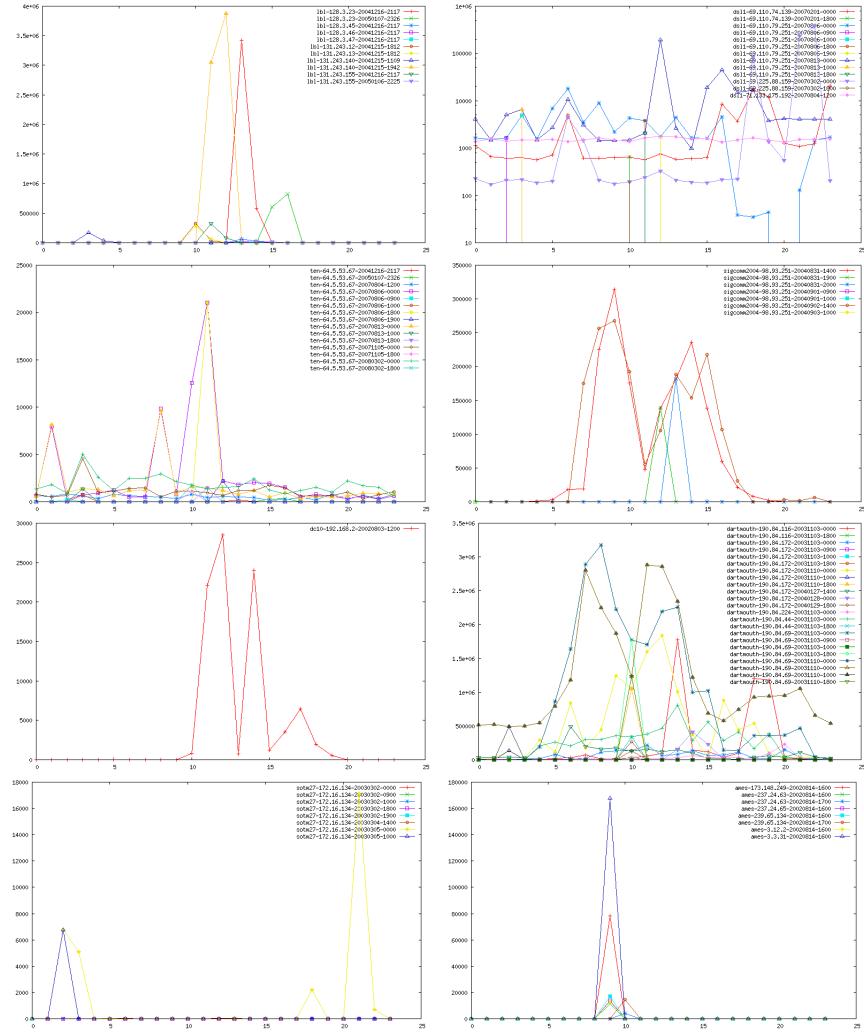


Figure K.14: Packet LocHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

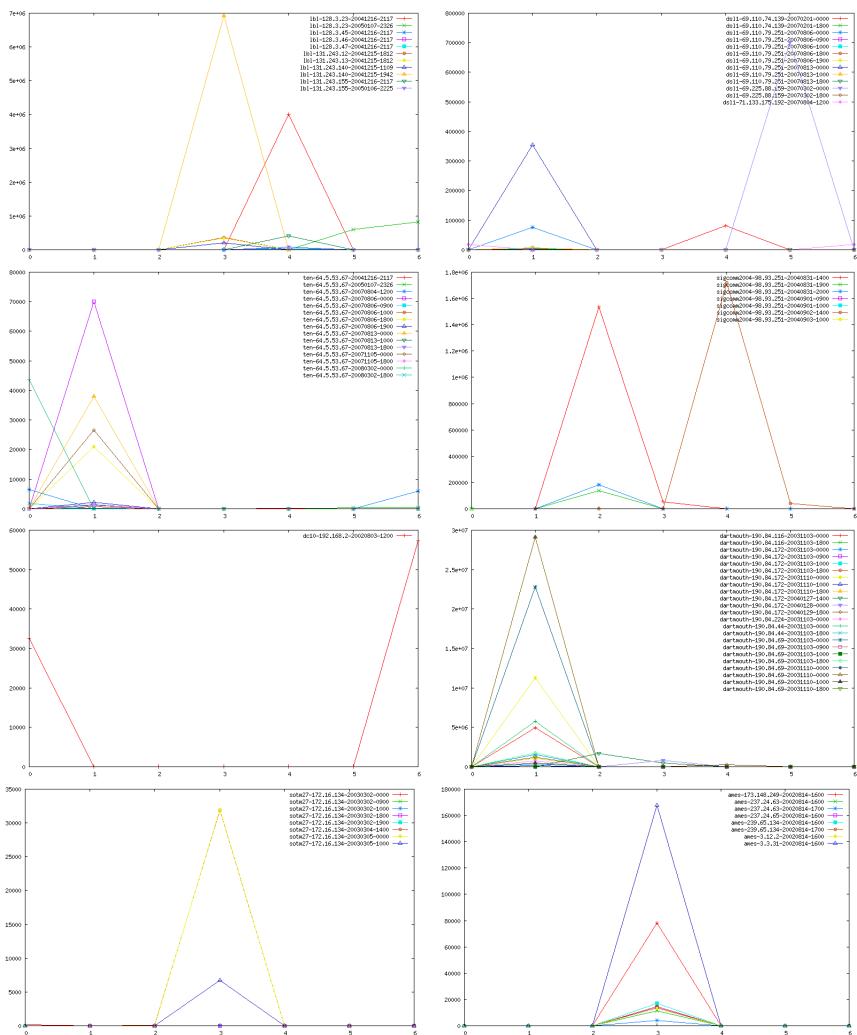


Figure K.15: Packet weekday distributions for all traces, with all the traces from the same dataset plotted together.



Figure K.16: Bytes sec distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.



Figure K.17: Bytes/min distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

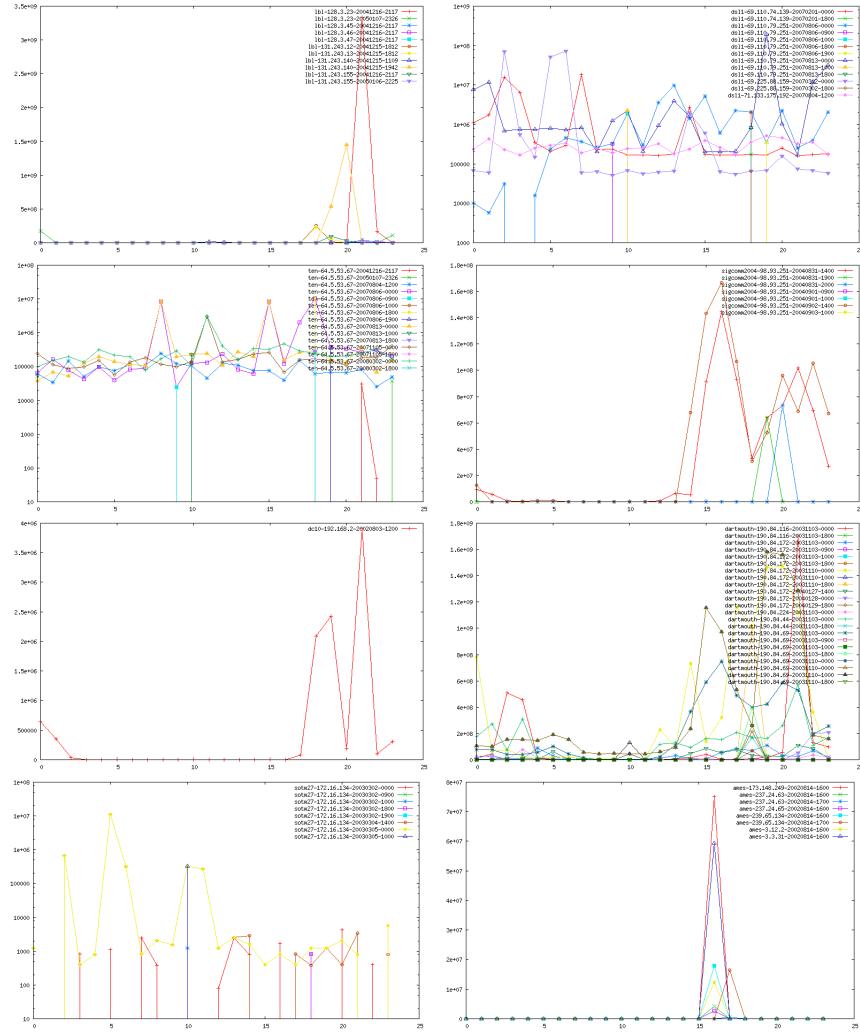


Figure K.18: Bytes GmHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

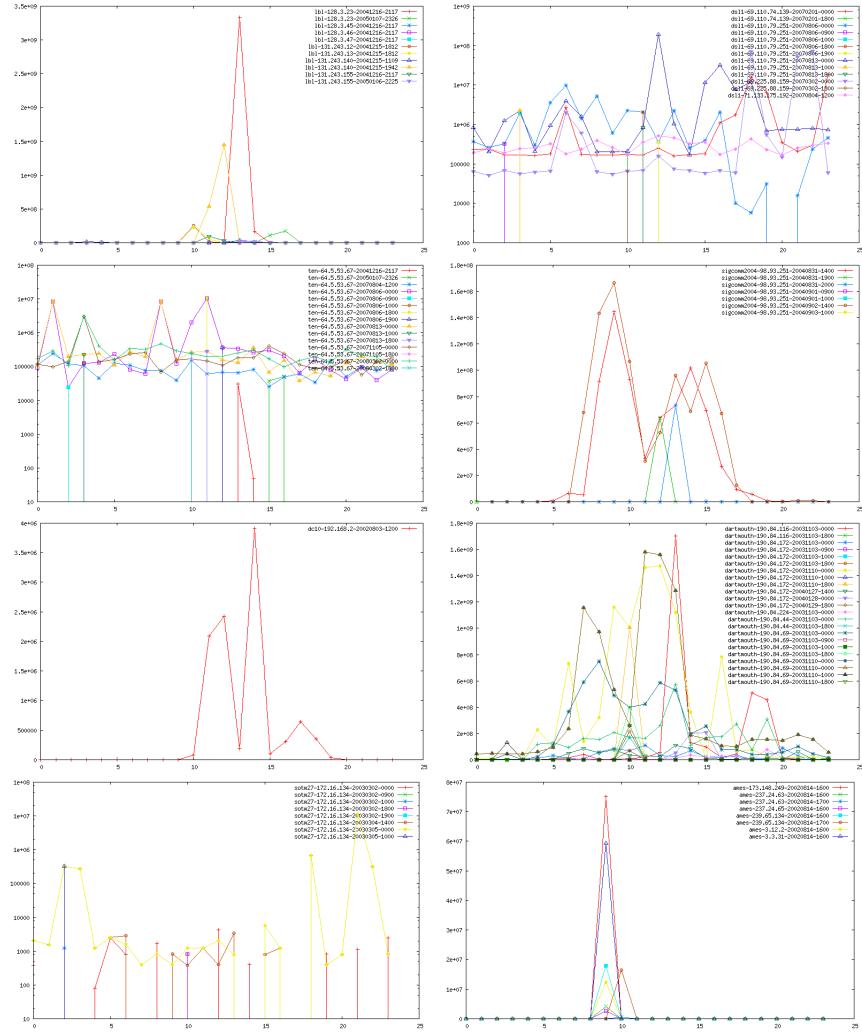


Figure K.19: Bytes LocHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

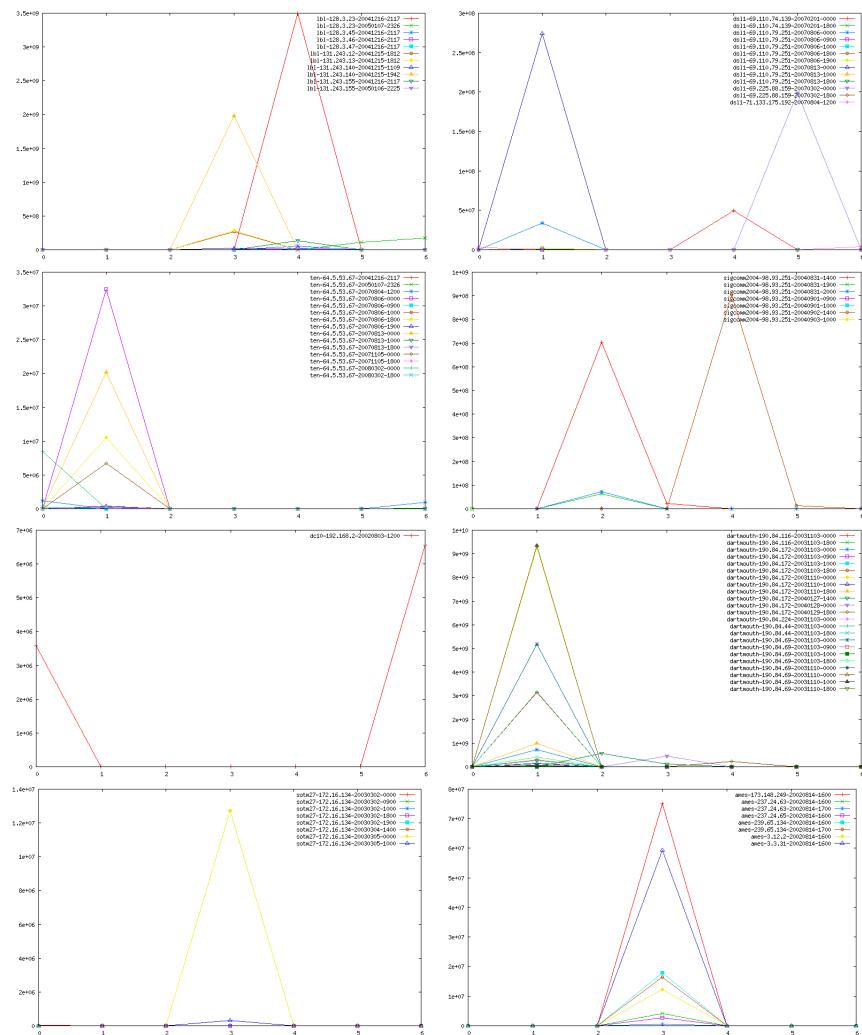


Figure K.20: Bytes weekday distributions for all traces, with all the traces from the same dataset plotted together.

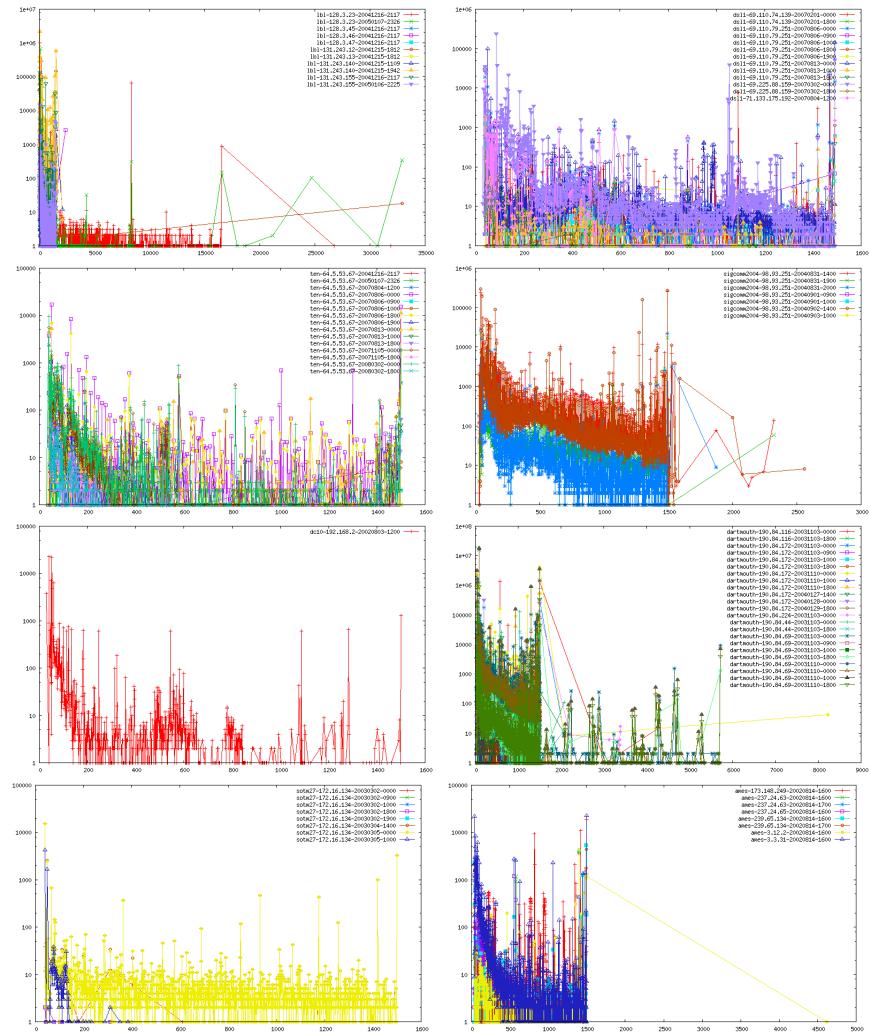


Figure K.21: Packet size distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

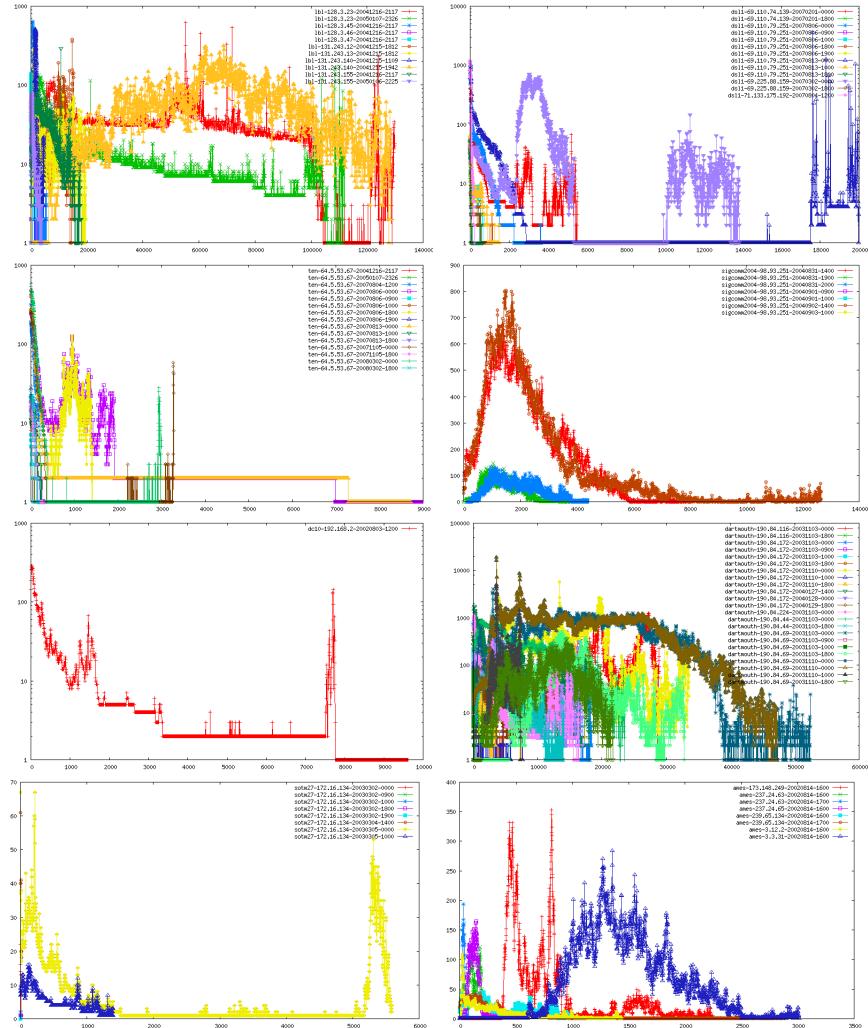


Figure K.22: Packets in last w secs distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

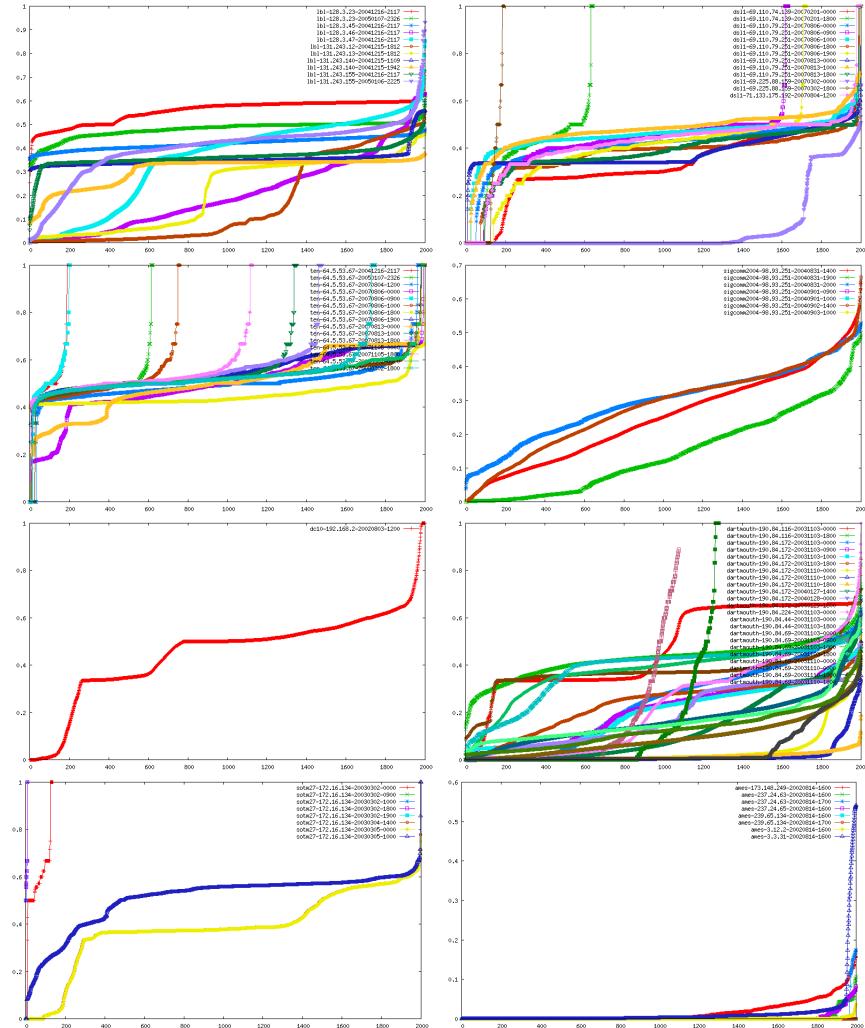


Figure K.23: Priv packets time rate distributions for all traces, with all the traces from the same dataset plotted together.

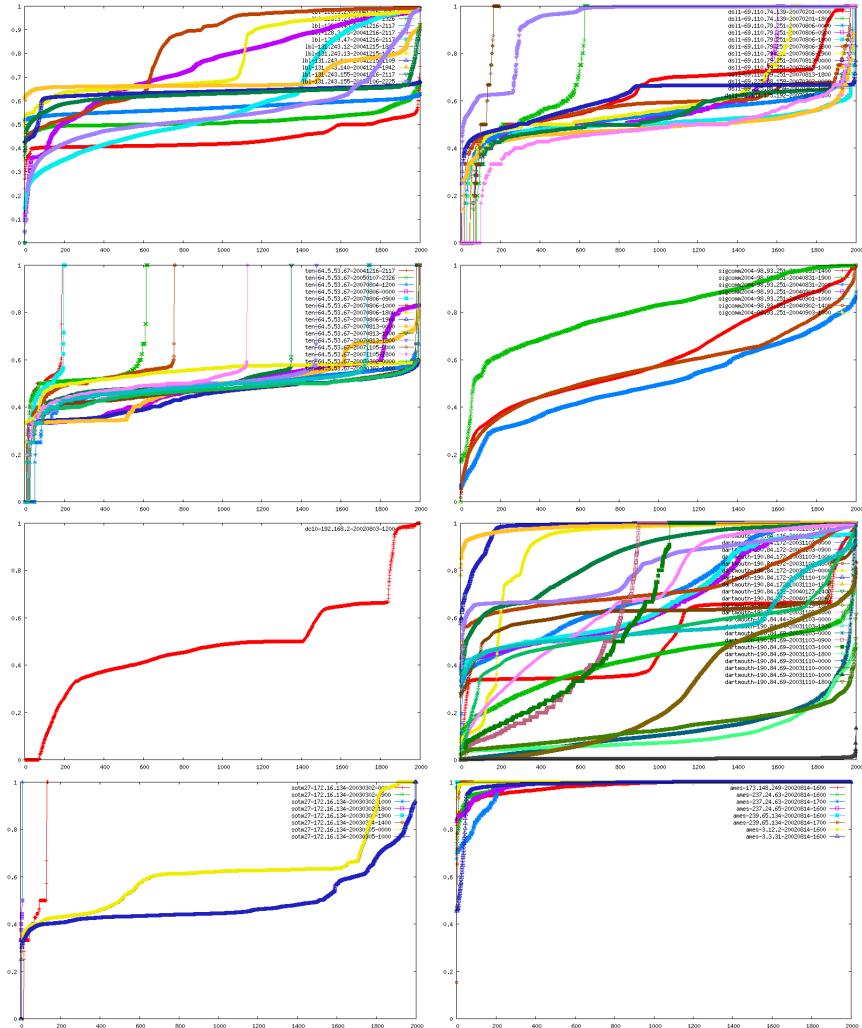


Figure K.24: Unpriv packets time rate distributions for all traces, with all the traces from the same dataset plotted together.

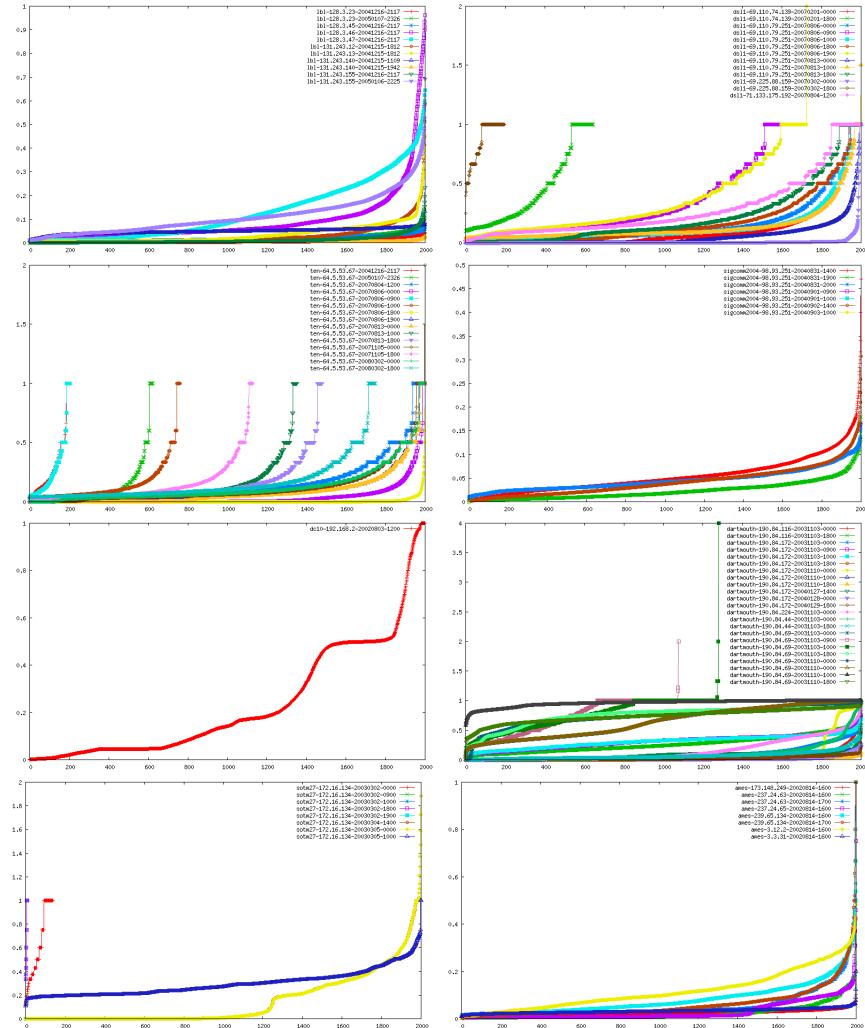


Figure K.25: Connections time rate distributions for all traces, with all the traces from the same dataset plotted together.

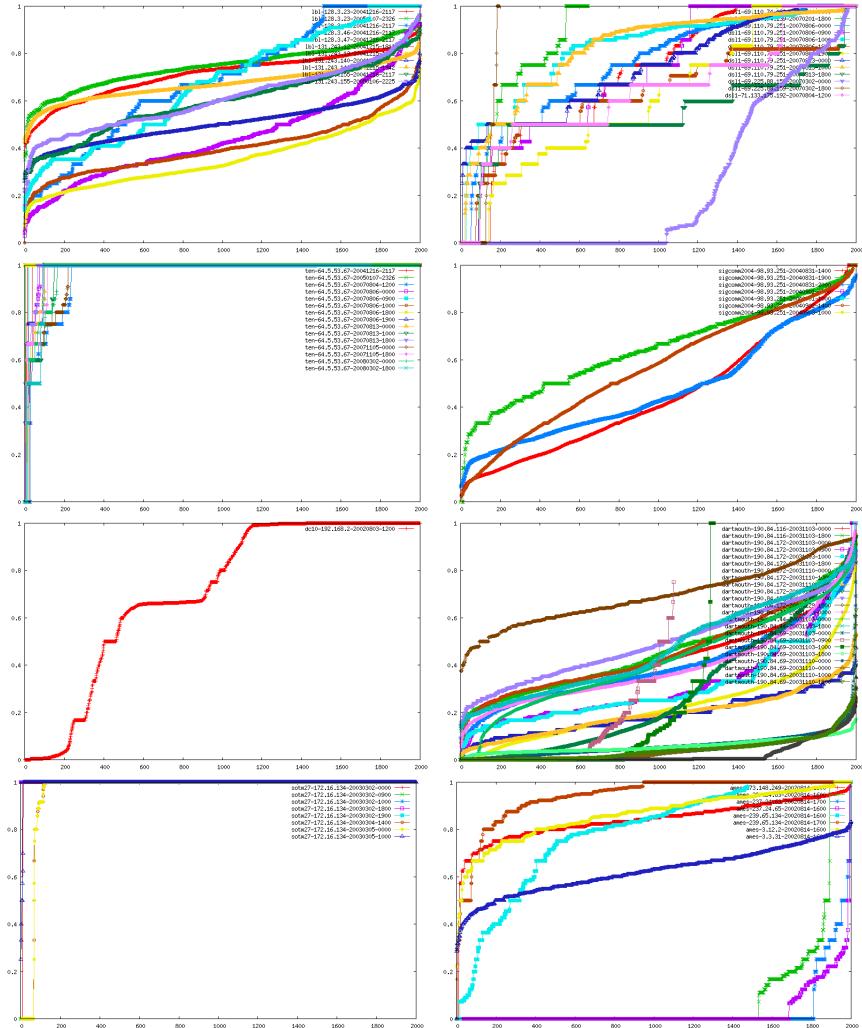


Figure K.26: Priv connections connection time rate distributions for all traces, with all the traces from the same dataset plotted together.

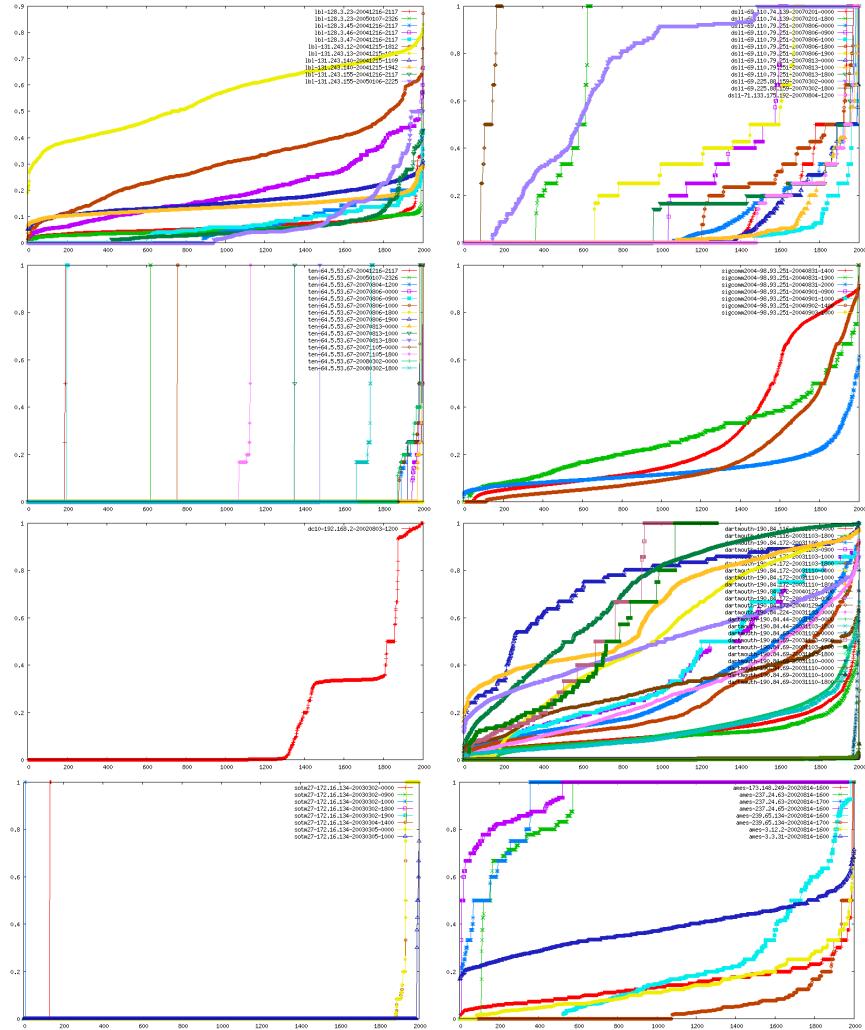


Figure K.27: Unpriv connections connection time rate distributions for all traces, with all the traces from the same dataset plotted together.

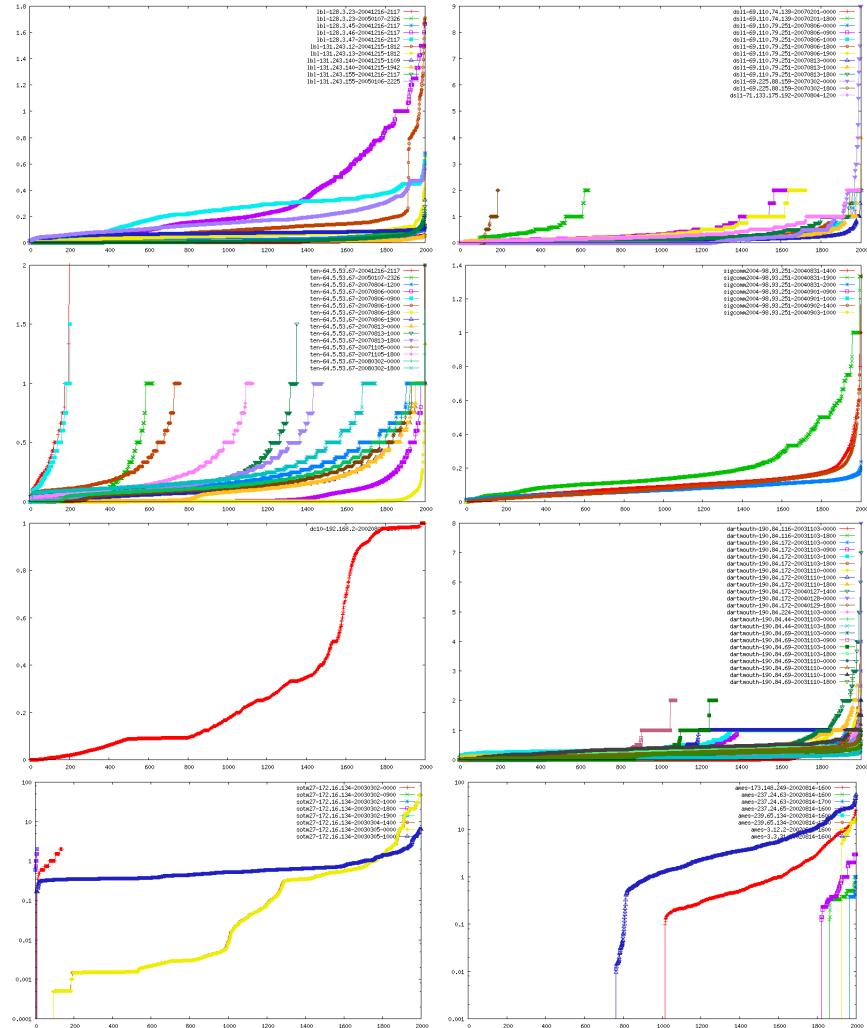


Figure K.28: Priv packets priv connection time rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

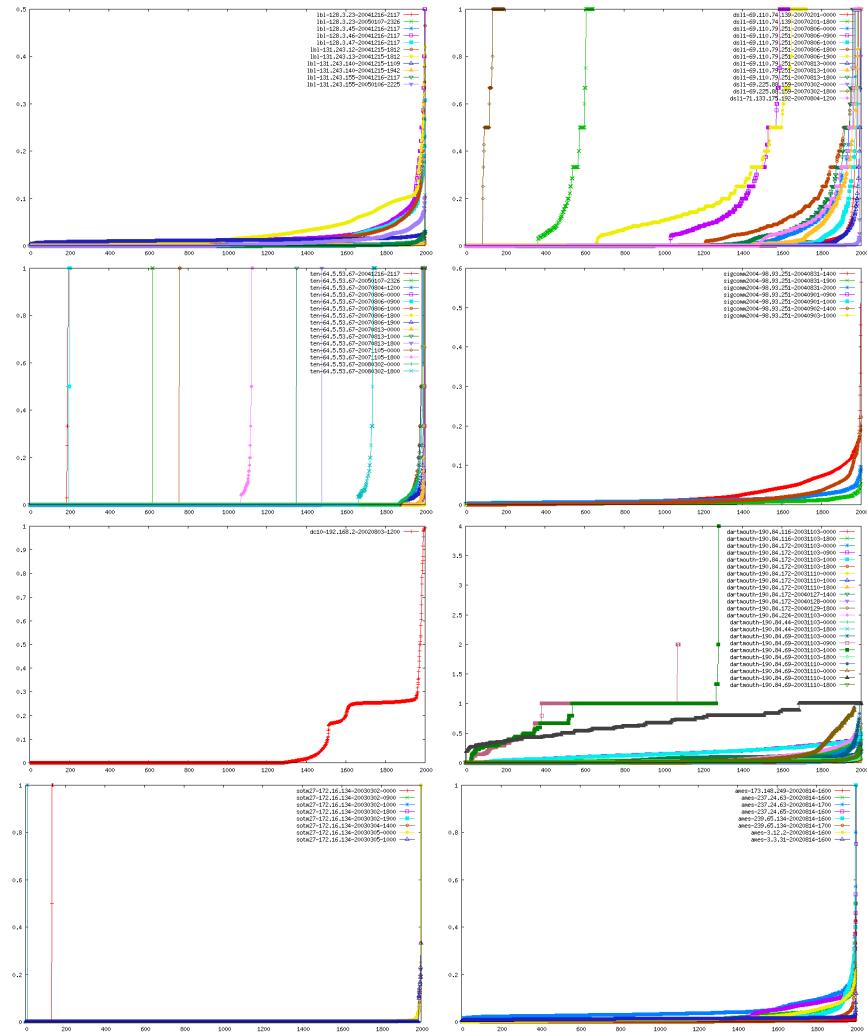


Figure K.29: Unpriv packets unpriv connection time rate distributions for all traces, with all the traces from the same dataset plotted together.

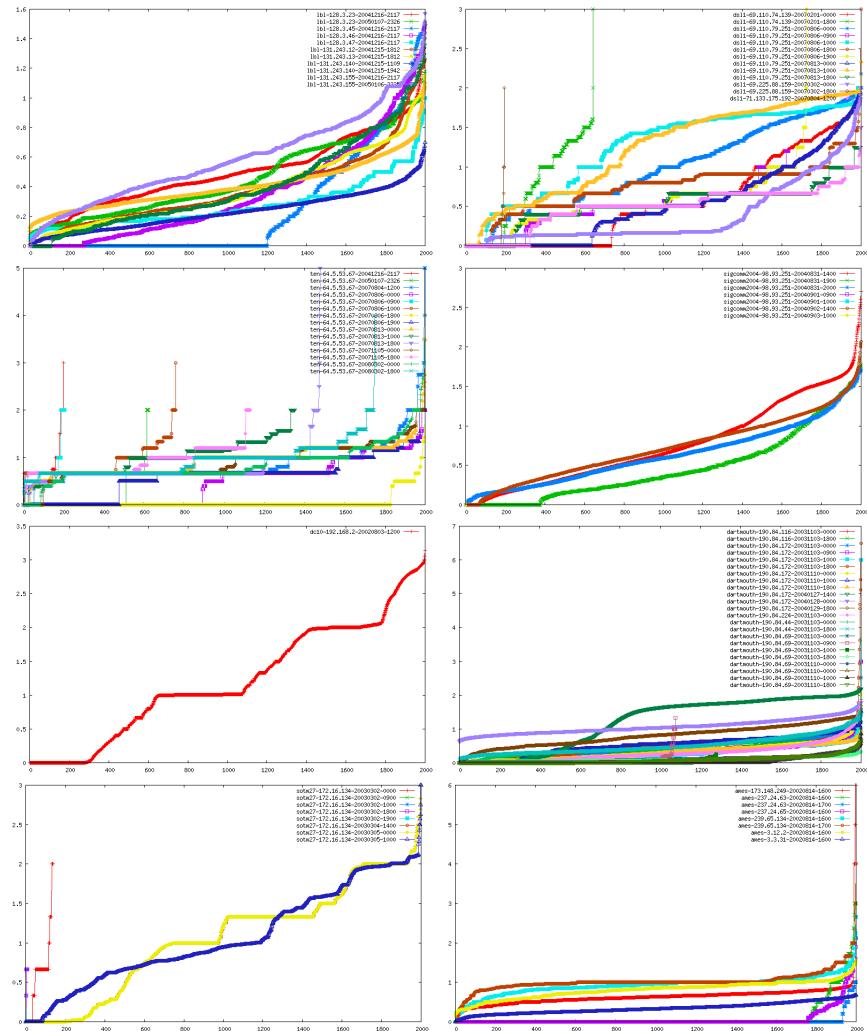


Figure K.30: SYNs connection time rate distributions for all traces, with all the traces from the same dataset plotted together.

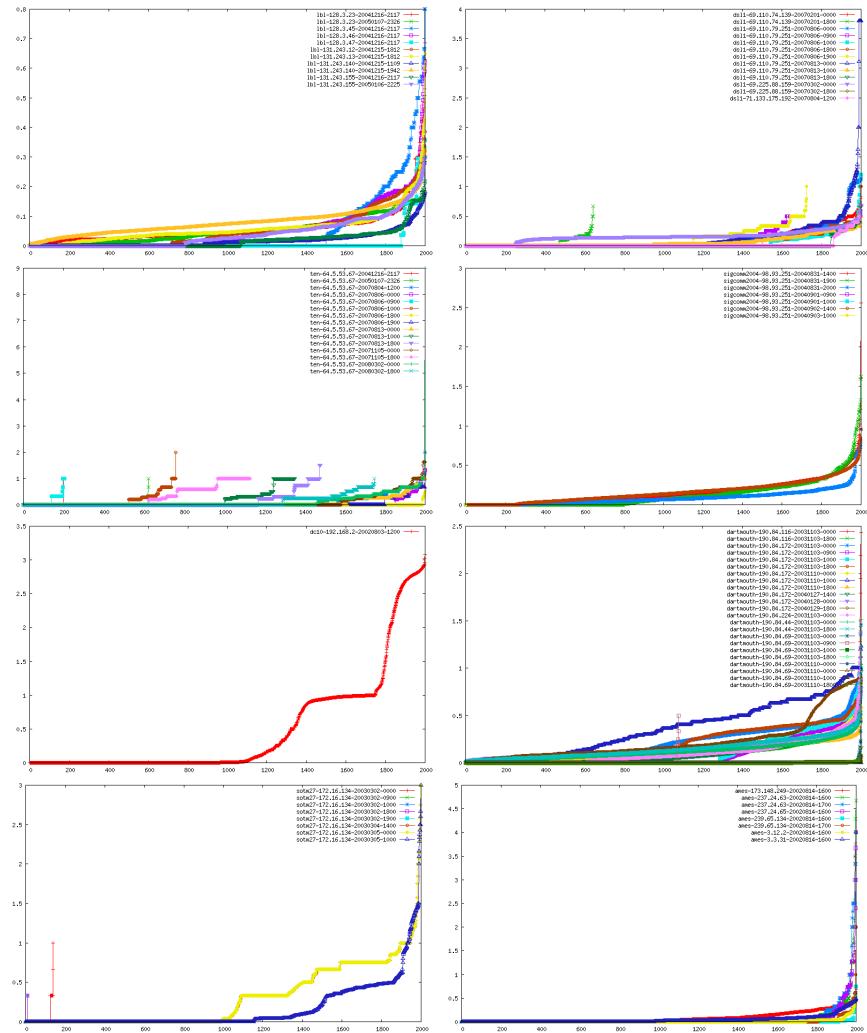


Figure K.31: RSTs connection time rate distributions for all traces, with all the traces from the same dataset plotted together.

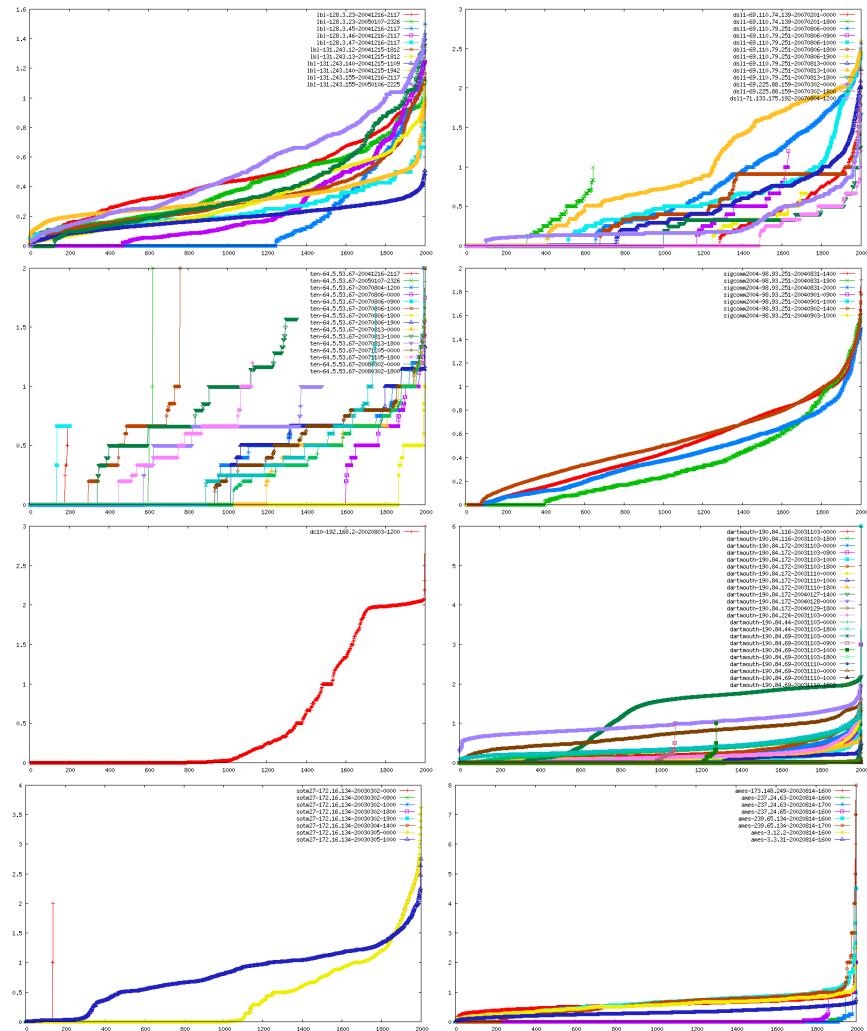


Figure K.32: FINs connection time rate distributions for all traces, with all the traces from the same dataset plotted together.

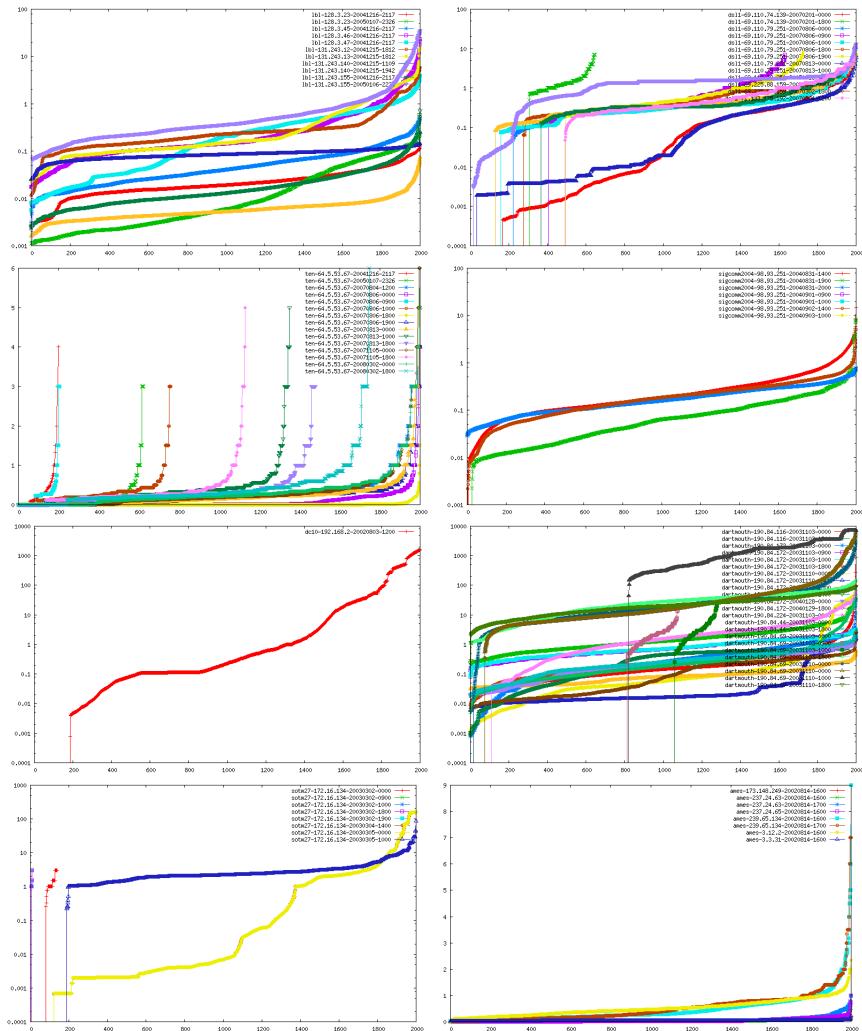


Figure K.33: PSH connection time rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

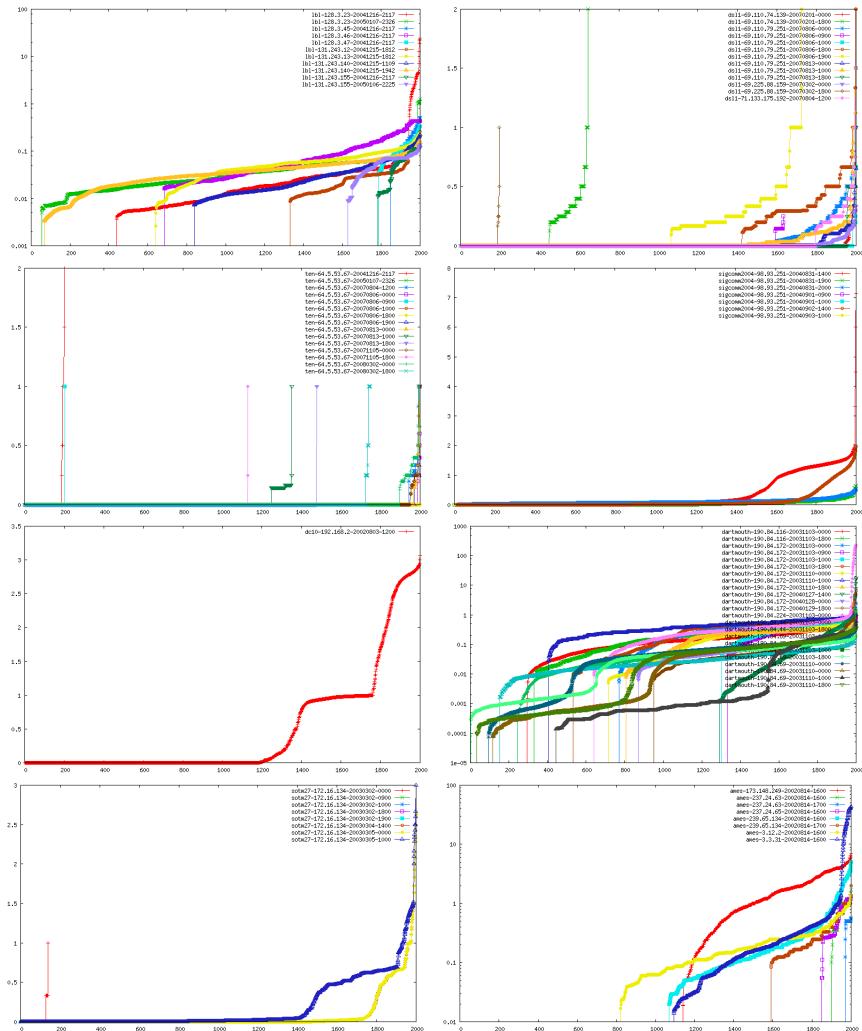


Figure K.34: Establishment errors connection time rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

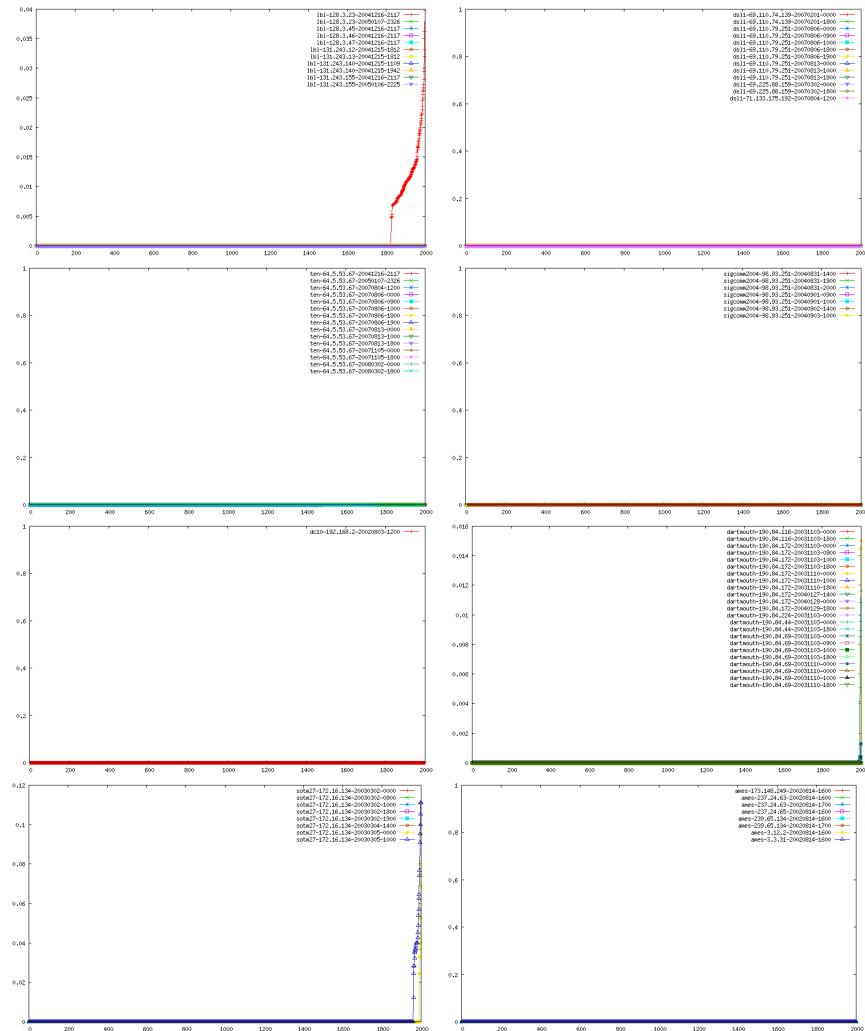


Figure K.35: Other errors connection time rate distributions for all traces, with all the traces from the same dataset plotted together.

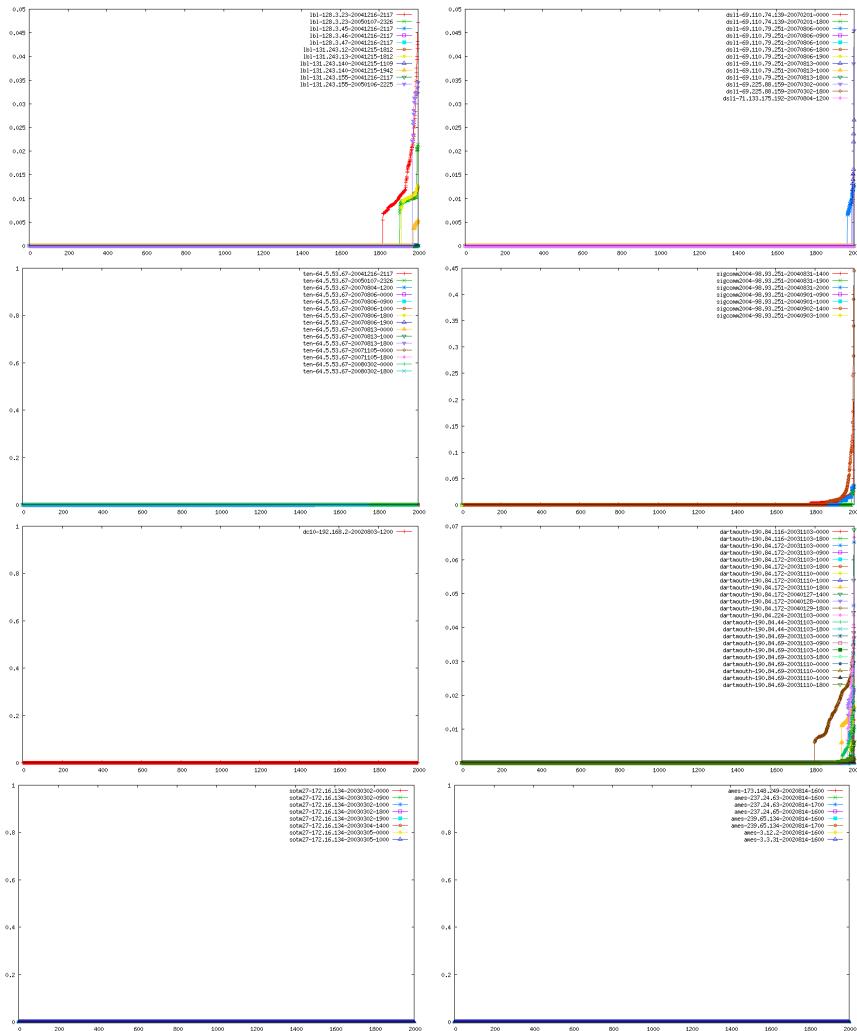


Figure K.36: Disconnection errors connection time rate distributions for all traces, with all the traces from the same dataset plotted together.

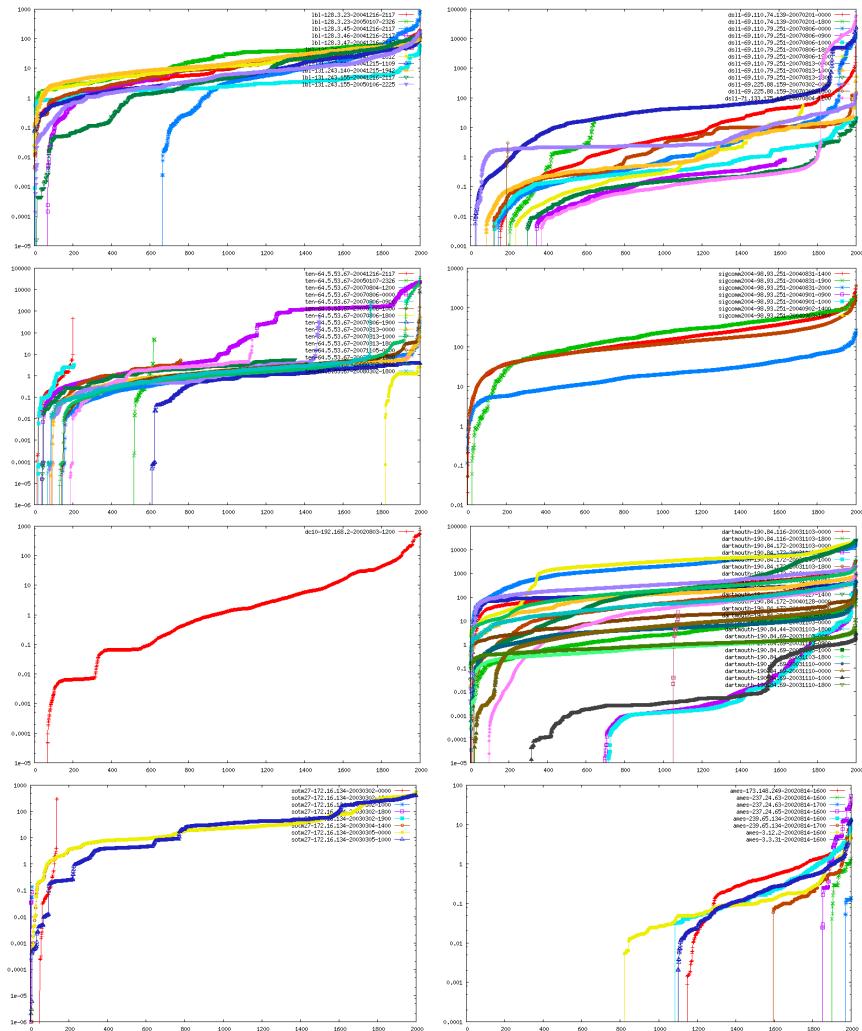


Figure K.37: Ave duration over last w secs distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

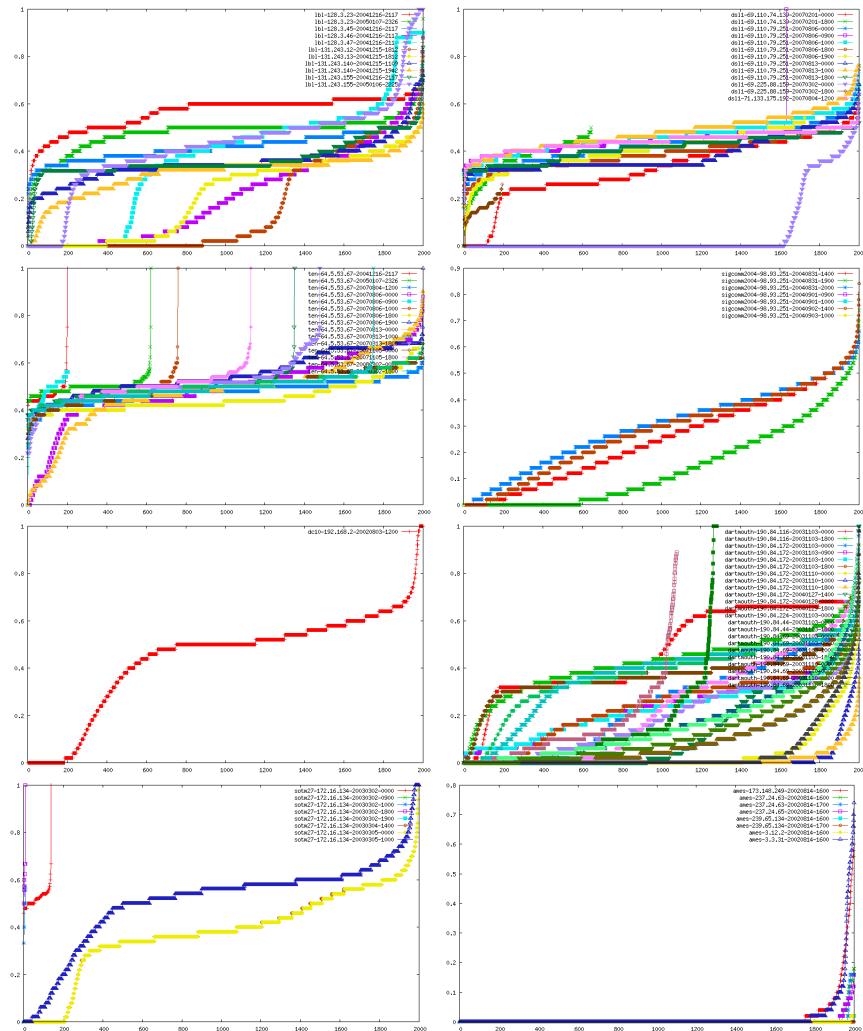


Figure K.38: Priv packets packet rate distributions for all traces, with all the traces from the same dataset plotted together.

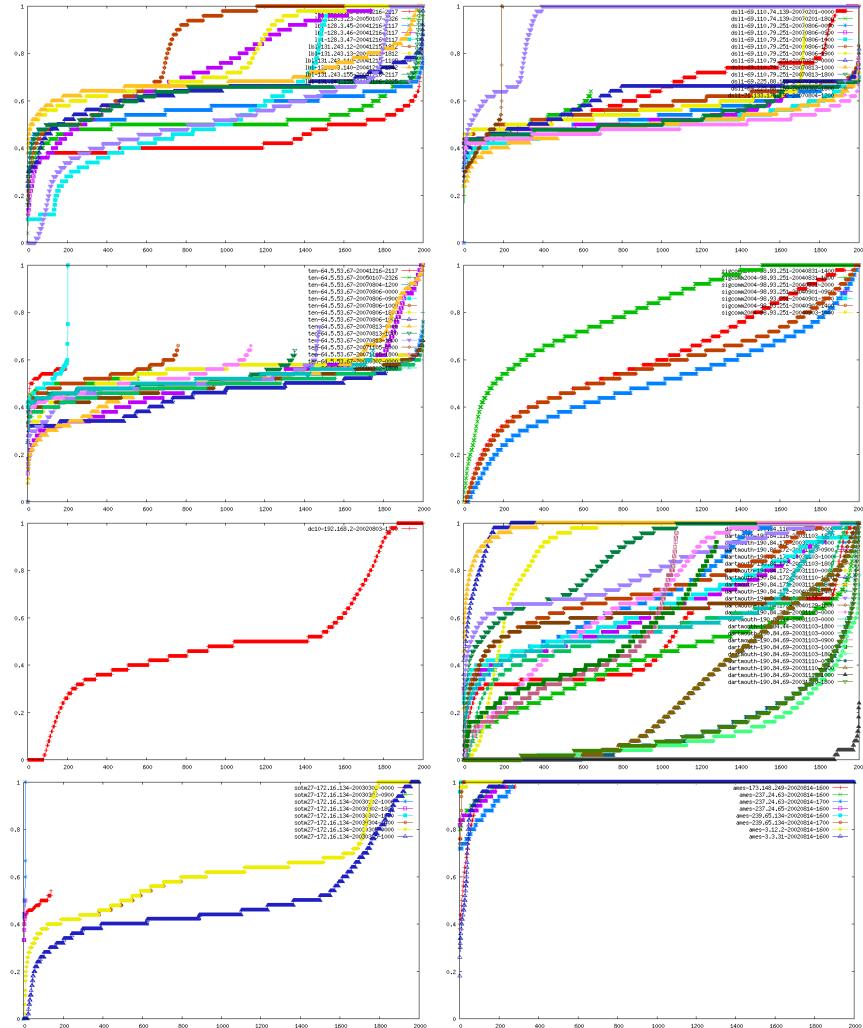


Figure K.39: Unpriv packets packet rate distributions for all traces, with all the traces from the same dataset plotted together.

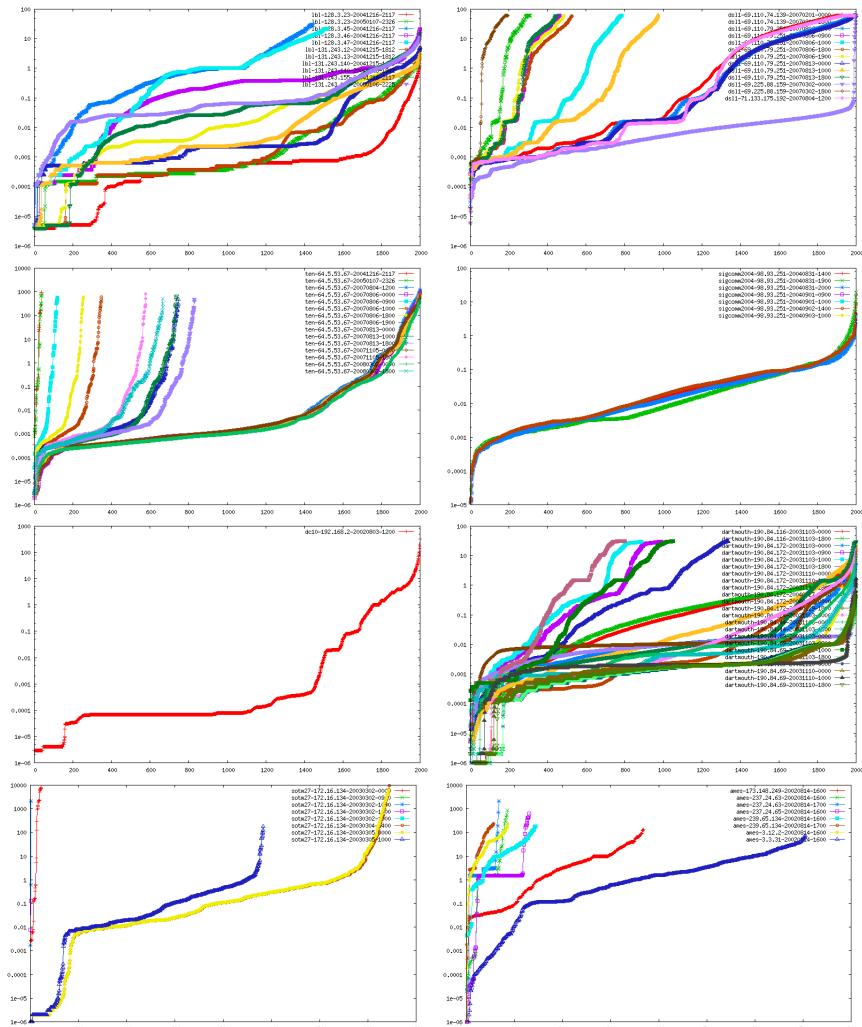


Figure K.40: InterConnection delta distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

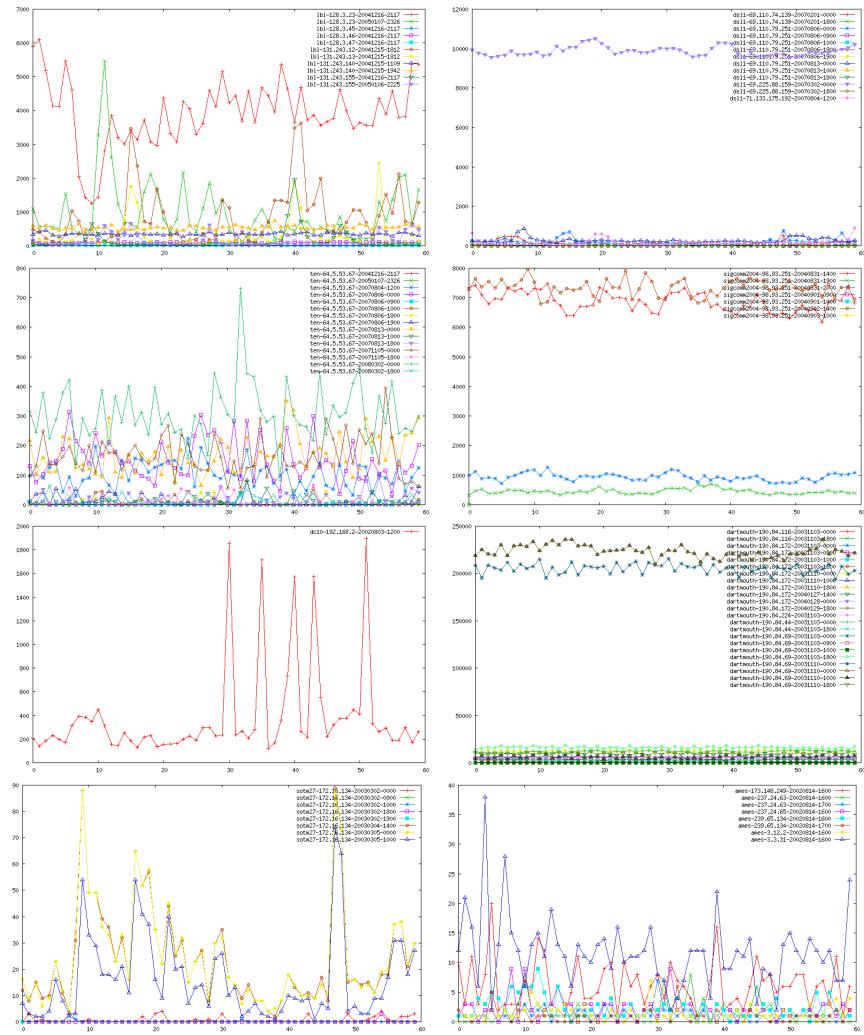


Figure K.41: Connection sec distributions for all traces, with all the traces from the same dataset plotted together.

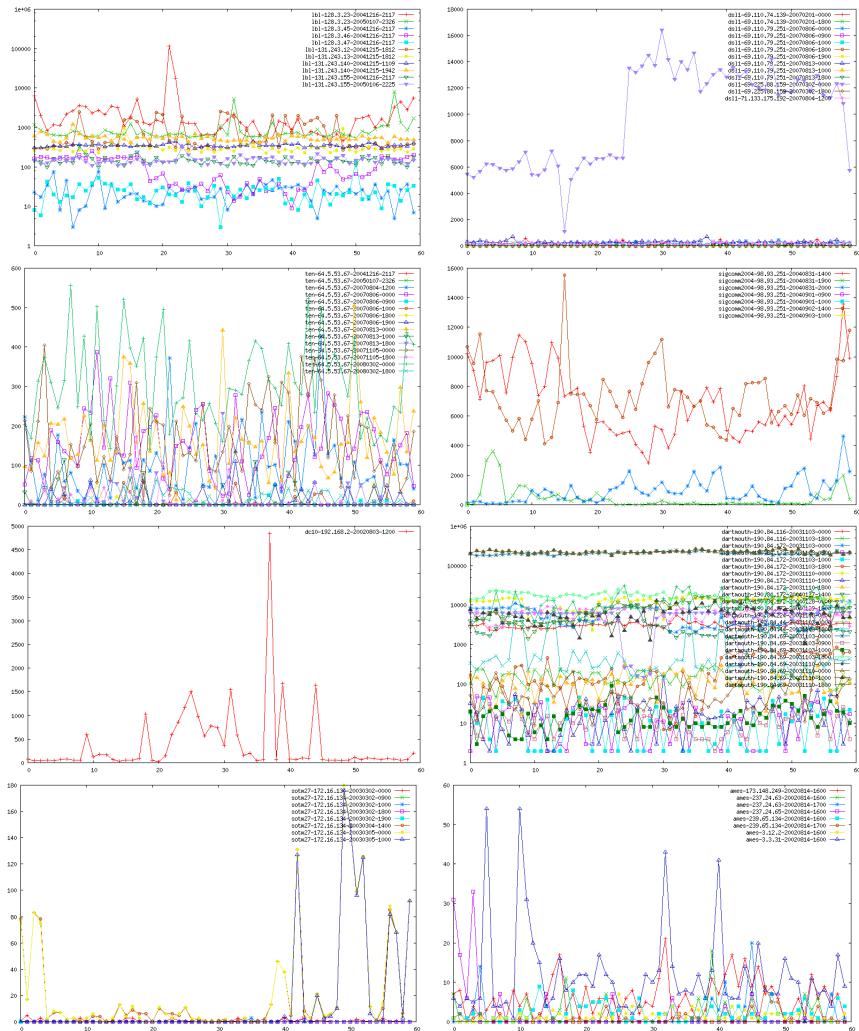


Figure K.42: Connection min distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

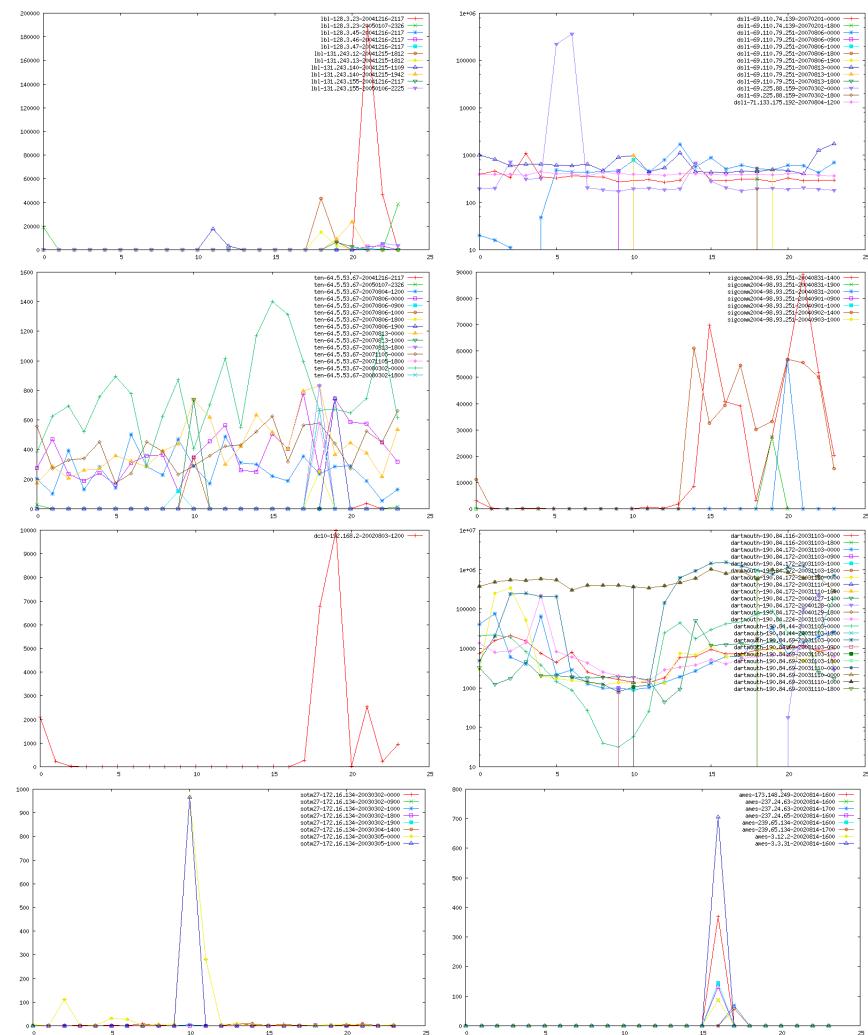


Figure K.43: Connection GmHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

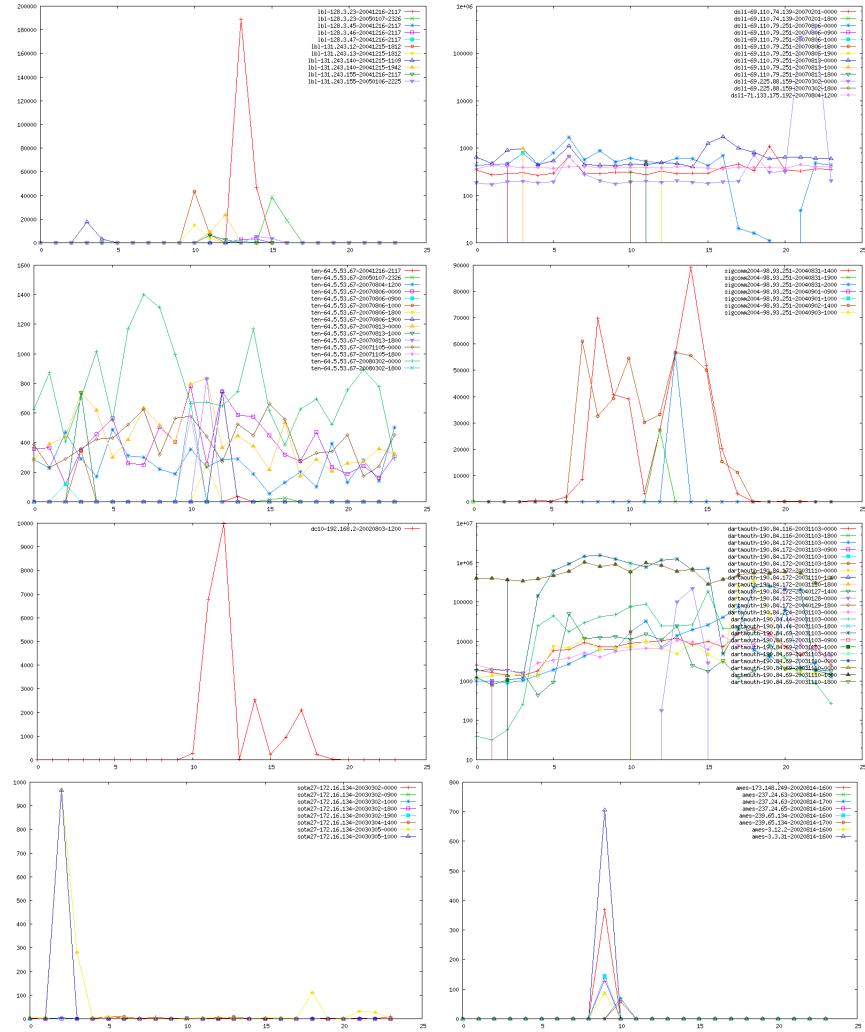


Figure K.44: Connection LocHour distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

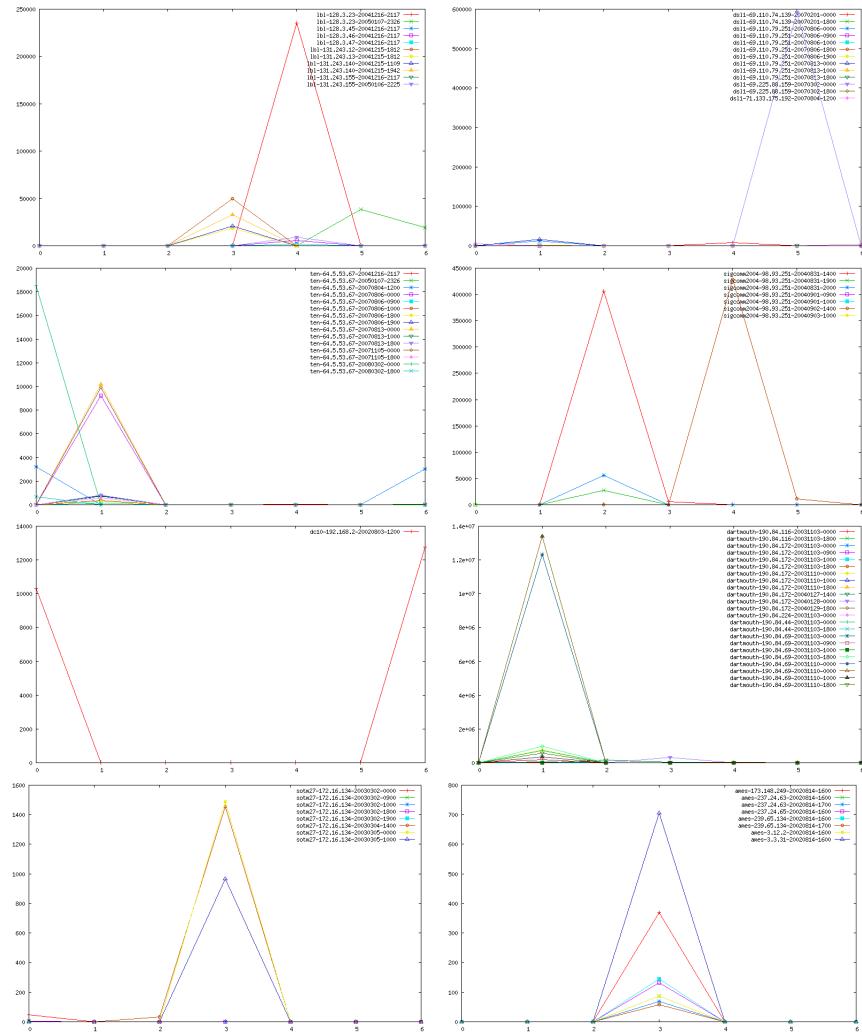


Figure K.45: Connection weekday distributions for all traces, with all the traces from the same dataset plotted together.

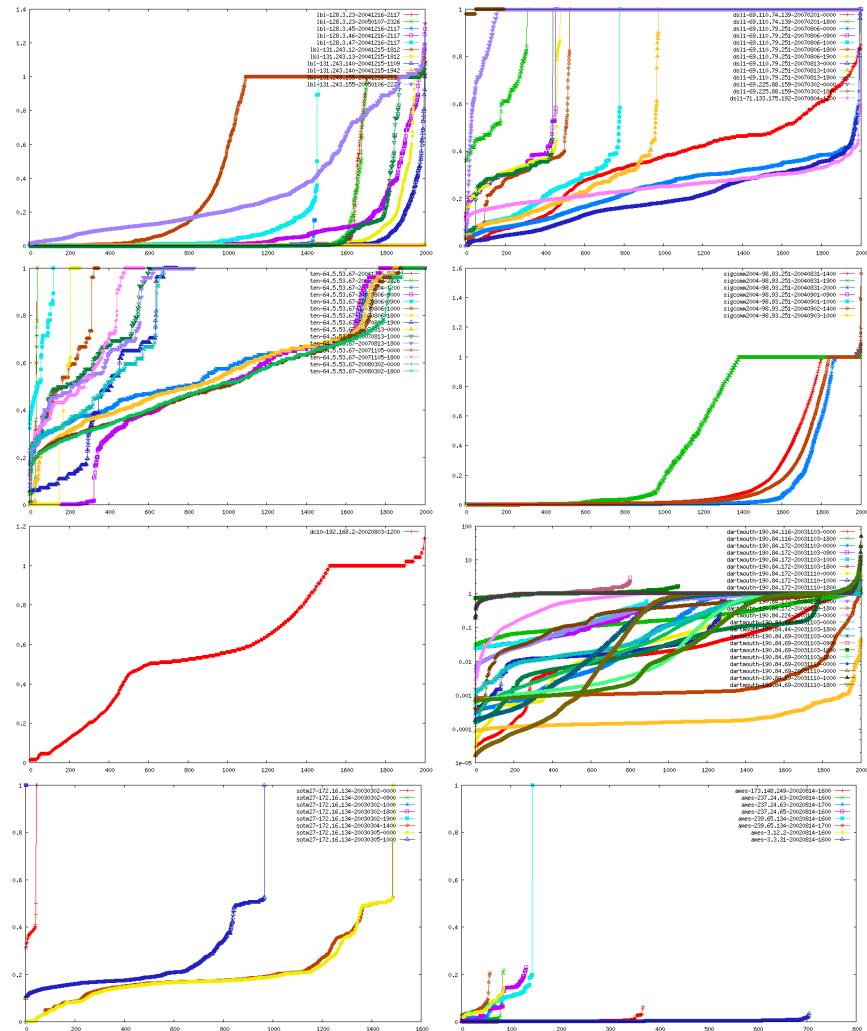


Figure K.46: Connection packet rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

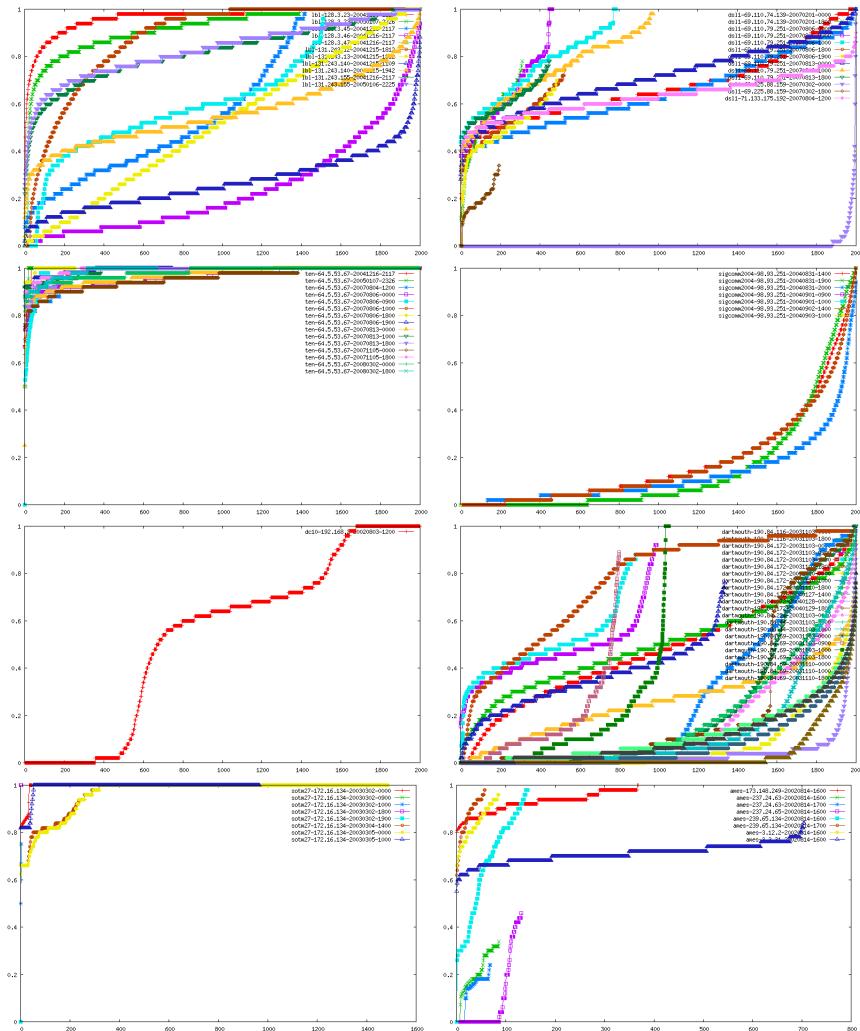


Figure K.47: Connection Priv connections rate distributions for all traces, with all the traces from the same dataset plotted together.

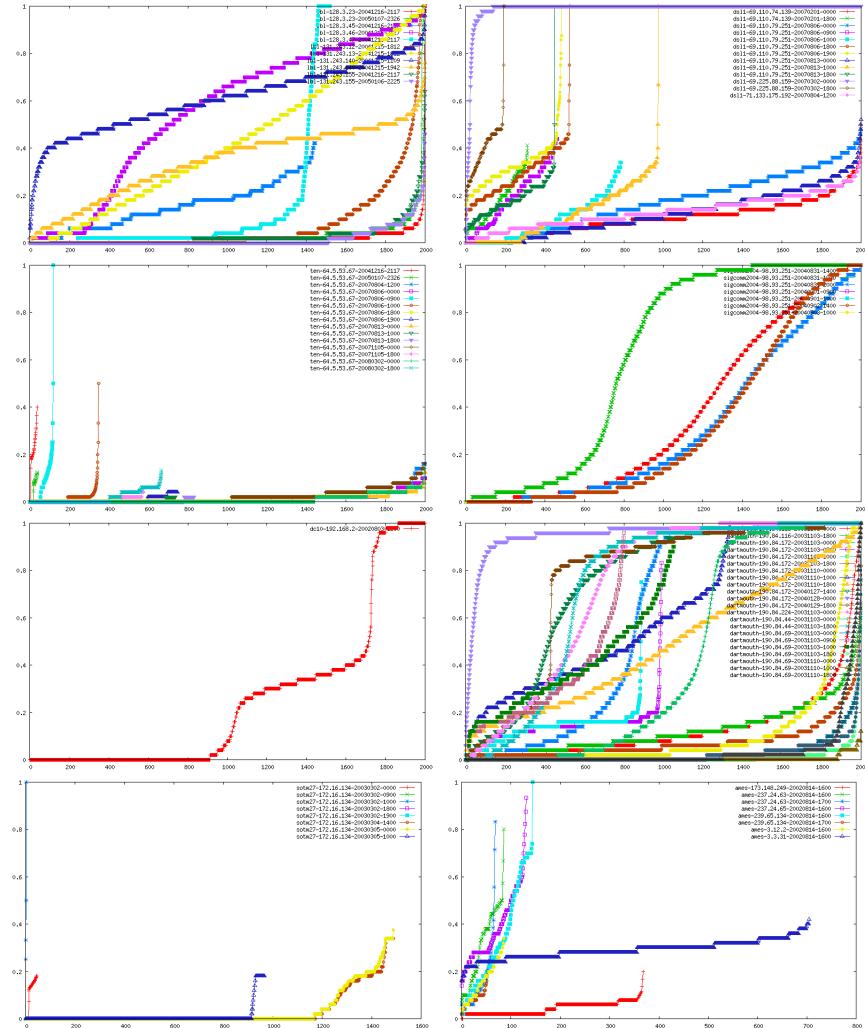


Figure K.48: Connection Unpriv connections rate distributions for all traces, with all the traces from the same dataset plotted together.

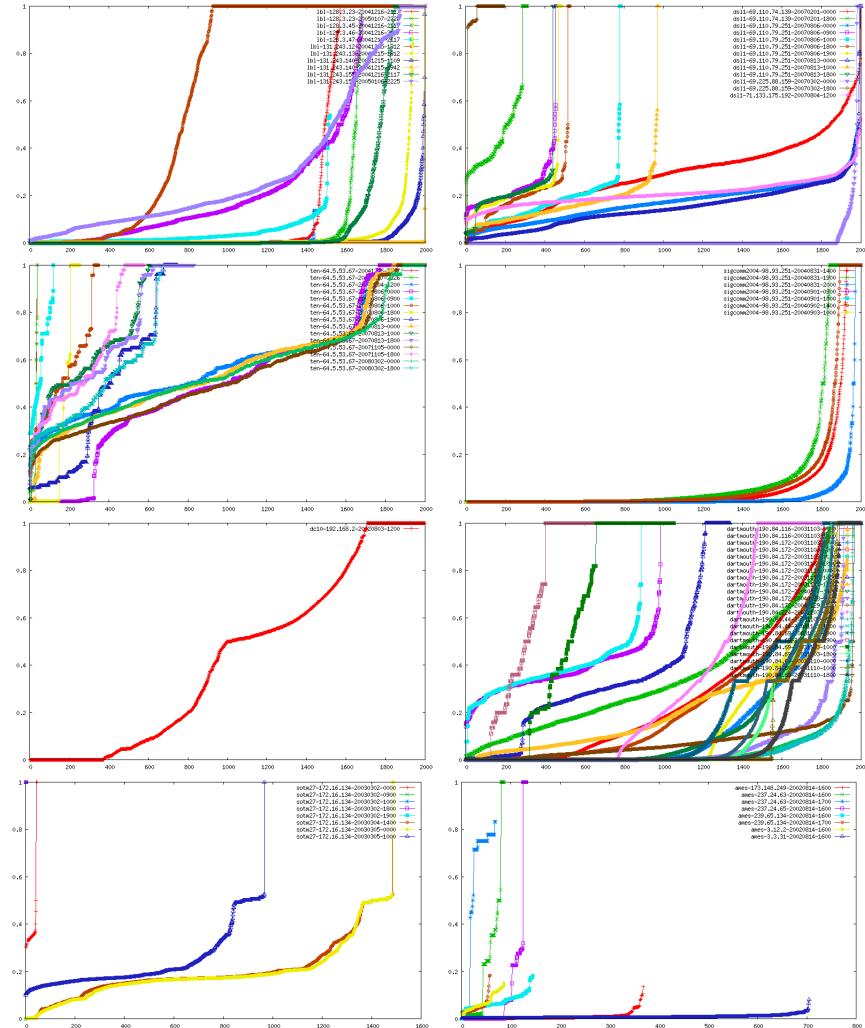


Figure K.49: Connection Priv packet rate distributions for all traces, with all the traces from the same dataset plotted together.

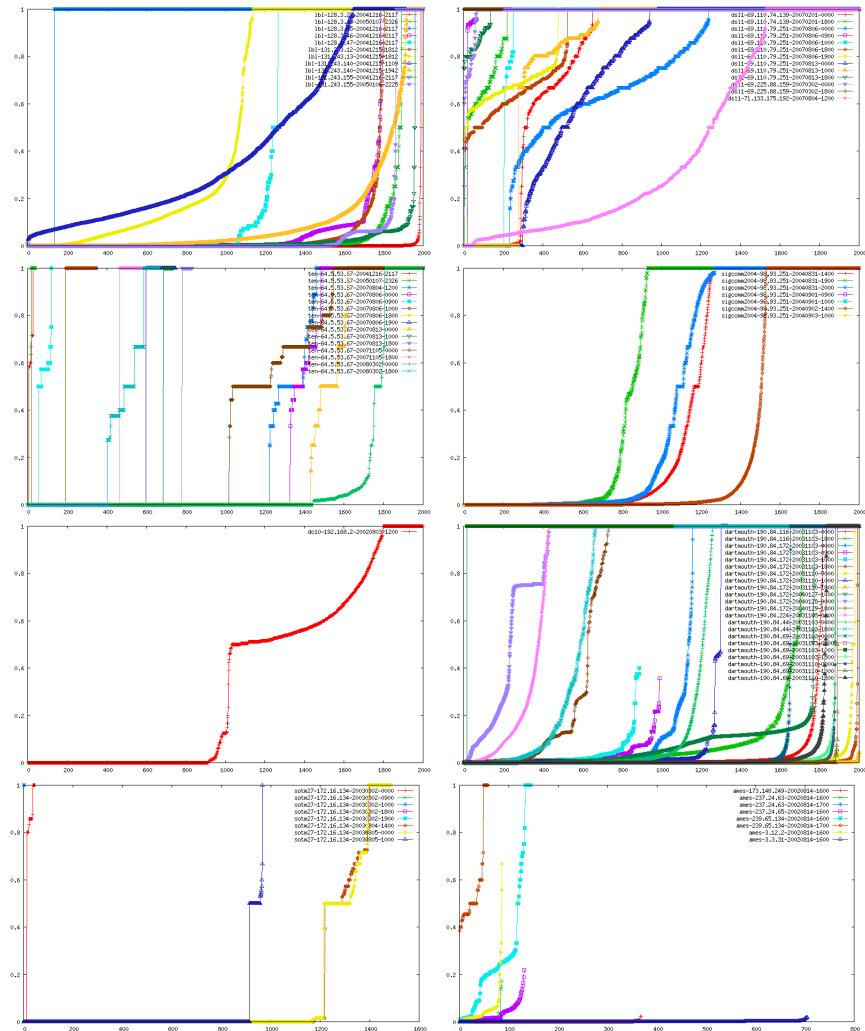


Figure K.50: Connection Unpriv packet rate distributions for all traces, with all the traces from the same dataset plotted together.

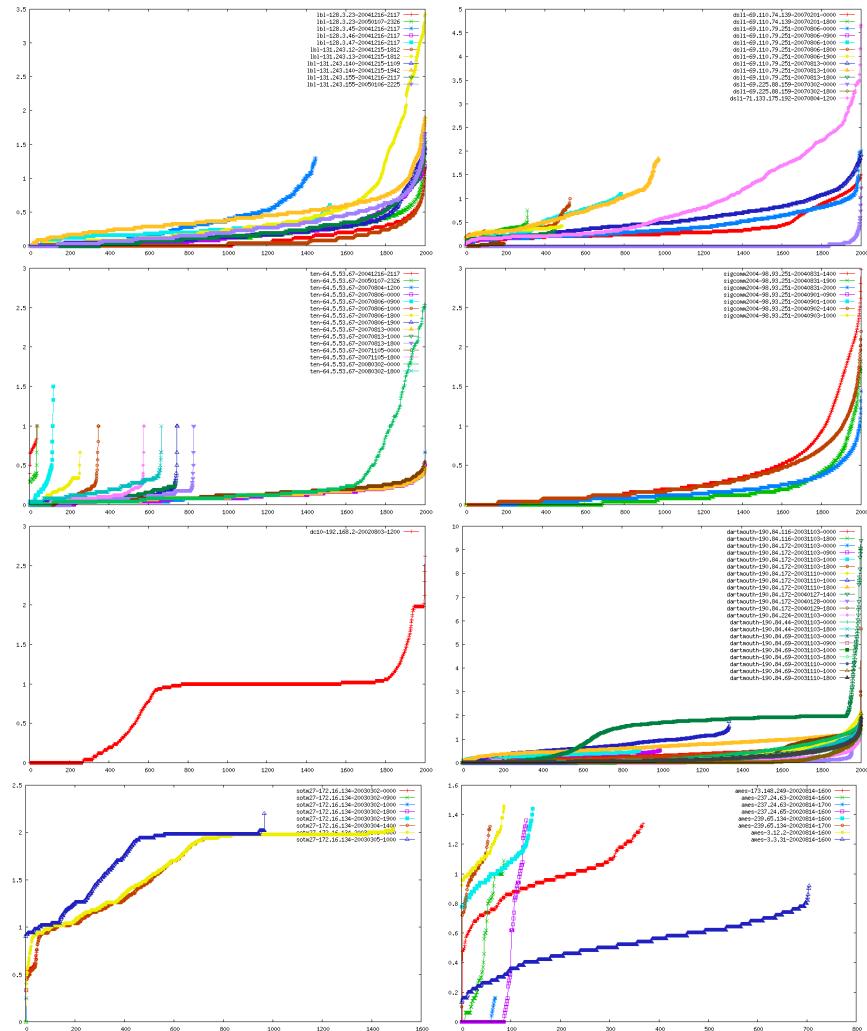


Figure K.51: Connection SYNs rate distributions for all traces, with all the traces from the same dataset plotted together.

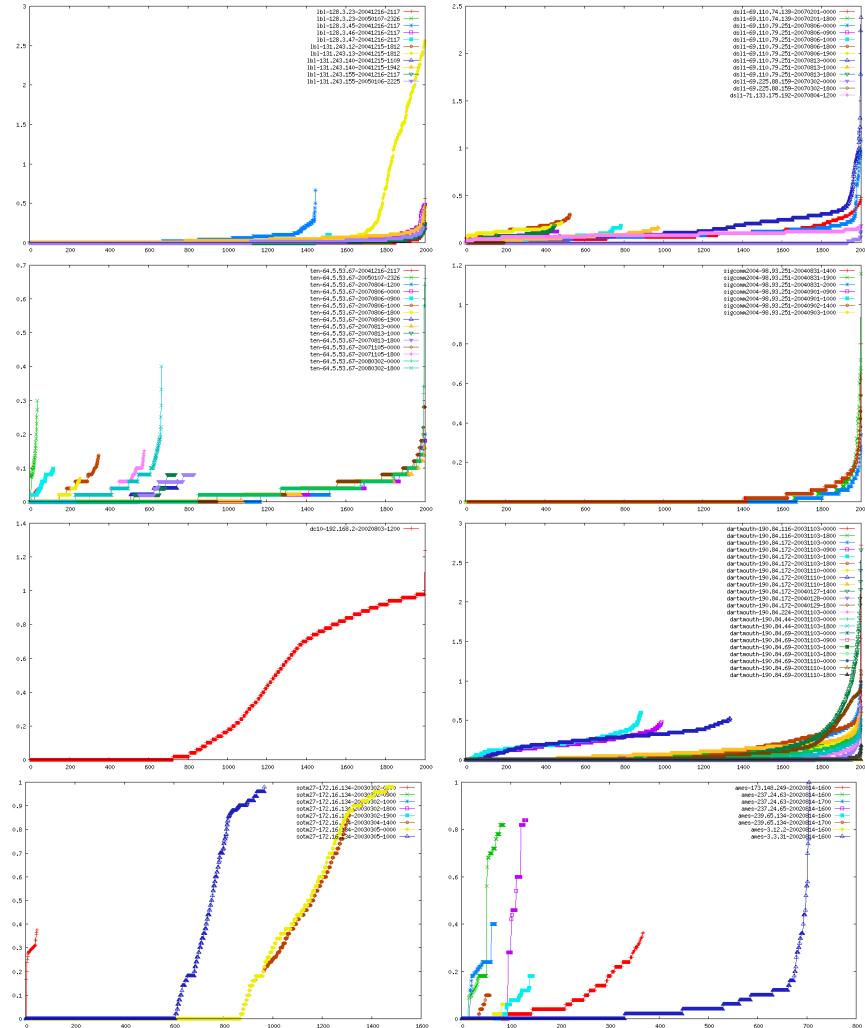


Figure K.52: Connection RSTs rate distributions for all traces, with all the traces from the same dataset plotted together.

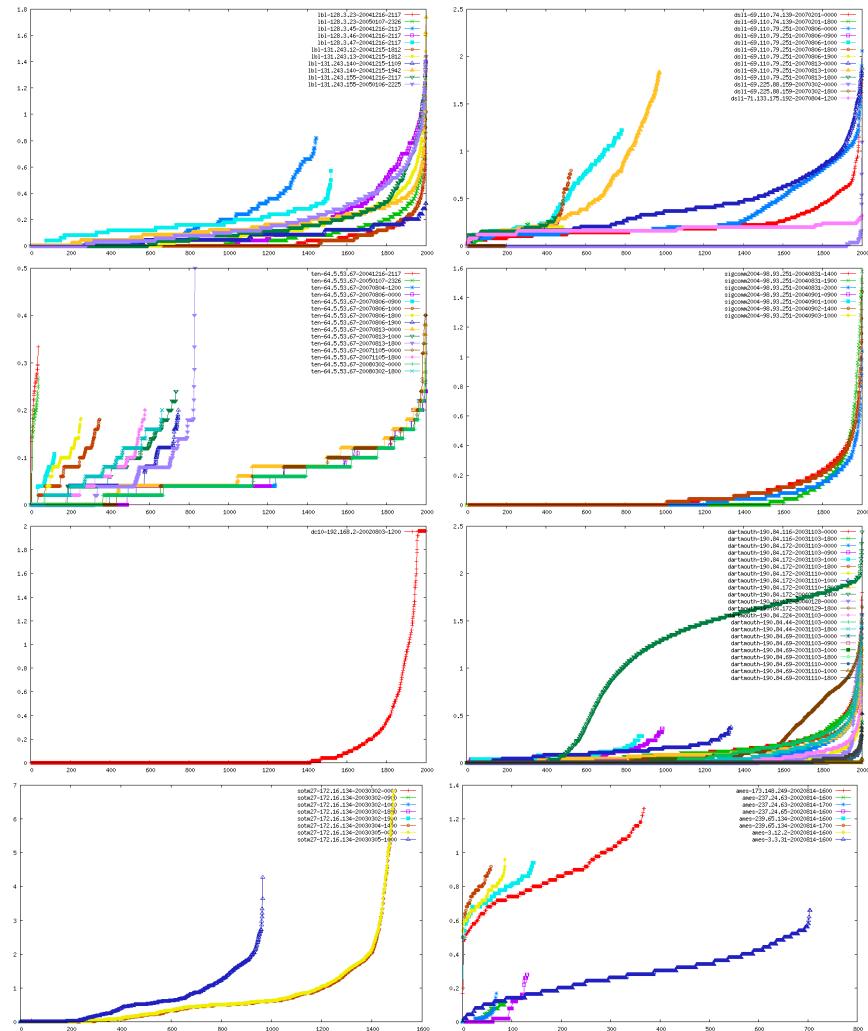


Figure K.53: Connection FINs rate distributions for all traces, with all the traces from the same dataset plotted together.

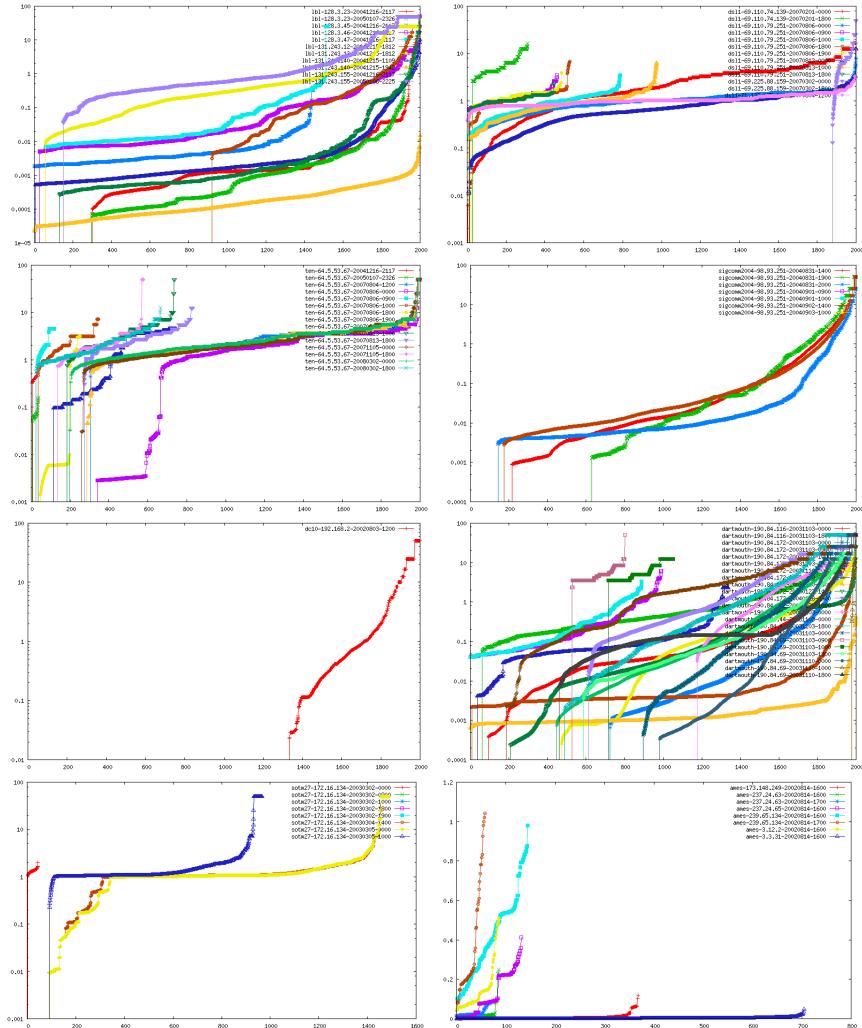


Figure K.54: Connection PSH rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

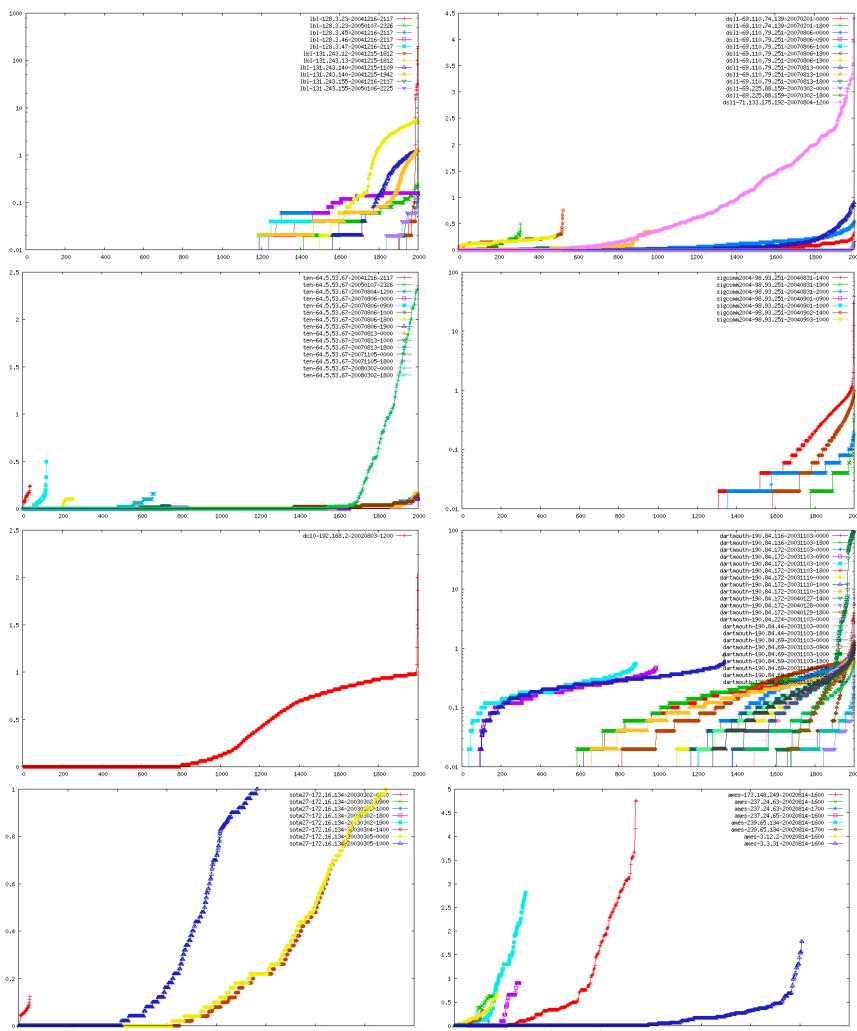


Figure K.55: Connection Establishment errors rate distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

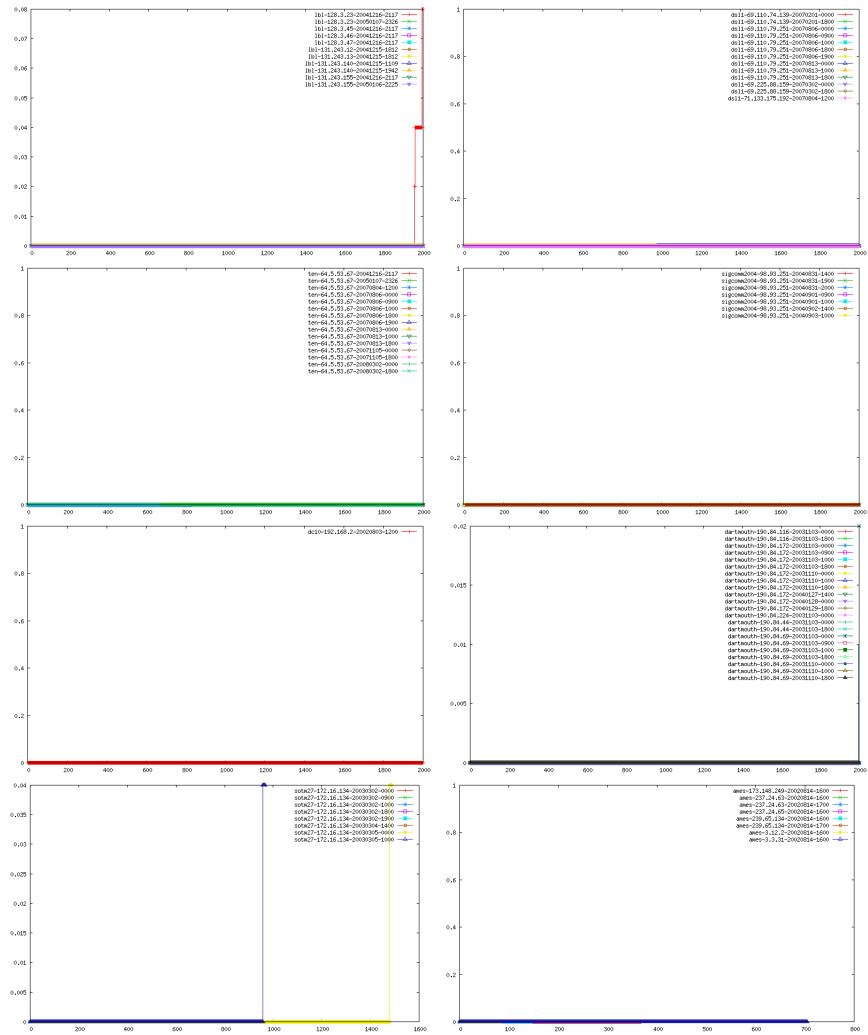


Figure K.56: Connection Other errors rate distributions for all traces, with all the traces from the same dataset plotted together.

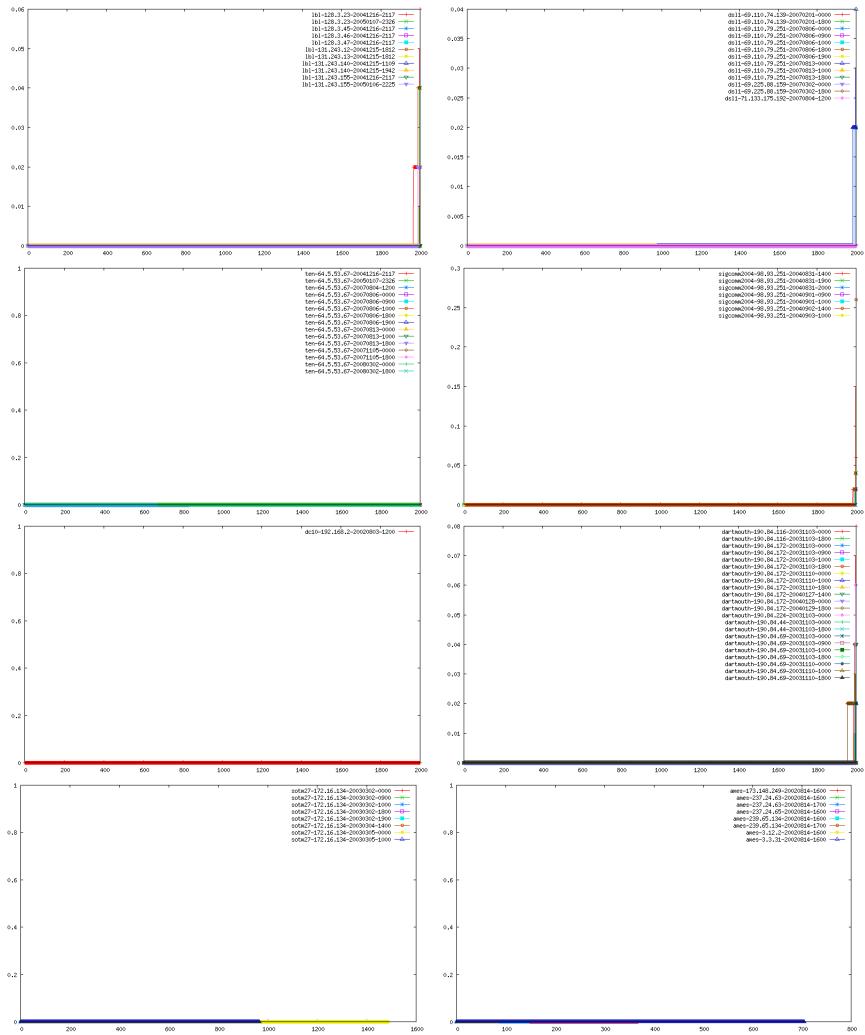


Figure K.57: Connection Disconnection errors rate distributions for all traces, with all the traces from the same dataset plotted together.

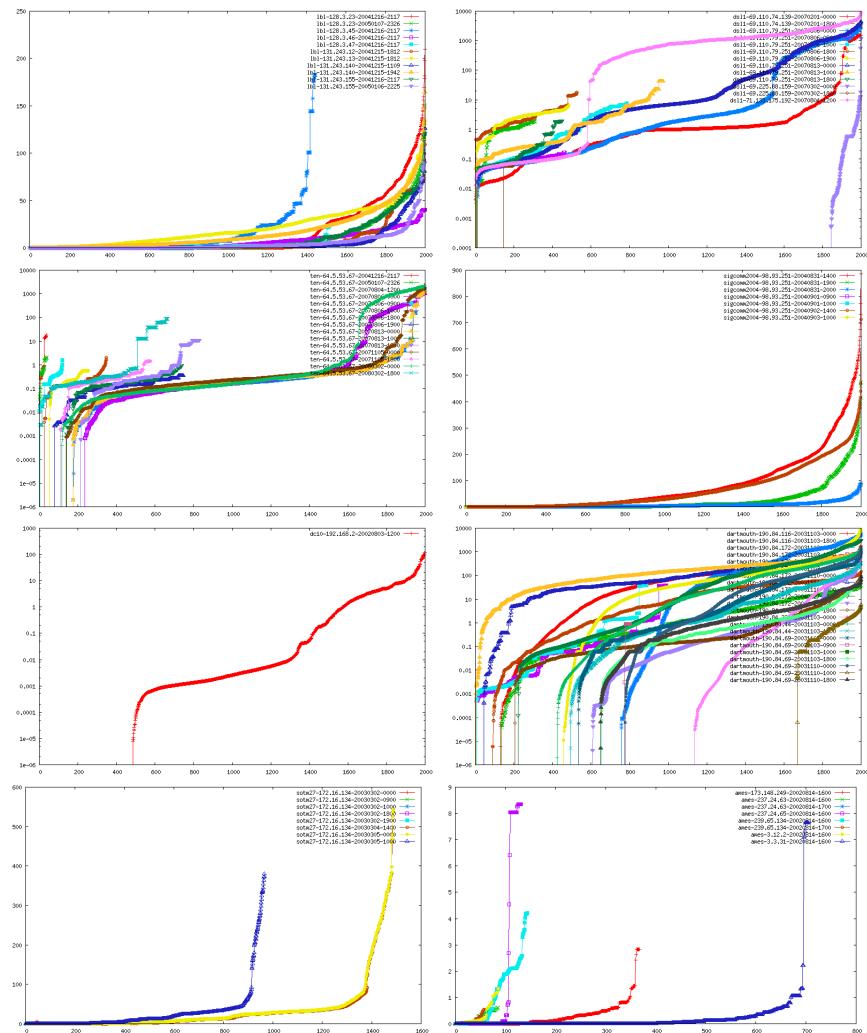


Figure K.58: Ave duration over last m connections distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

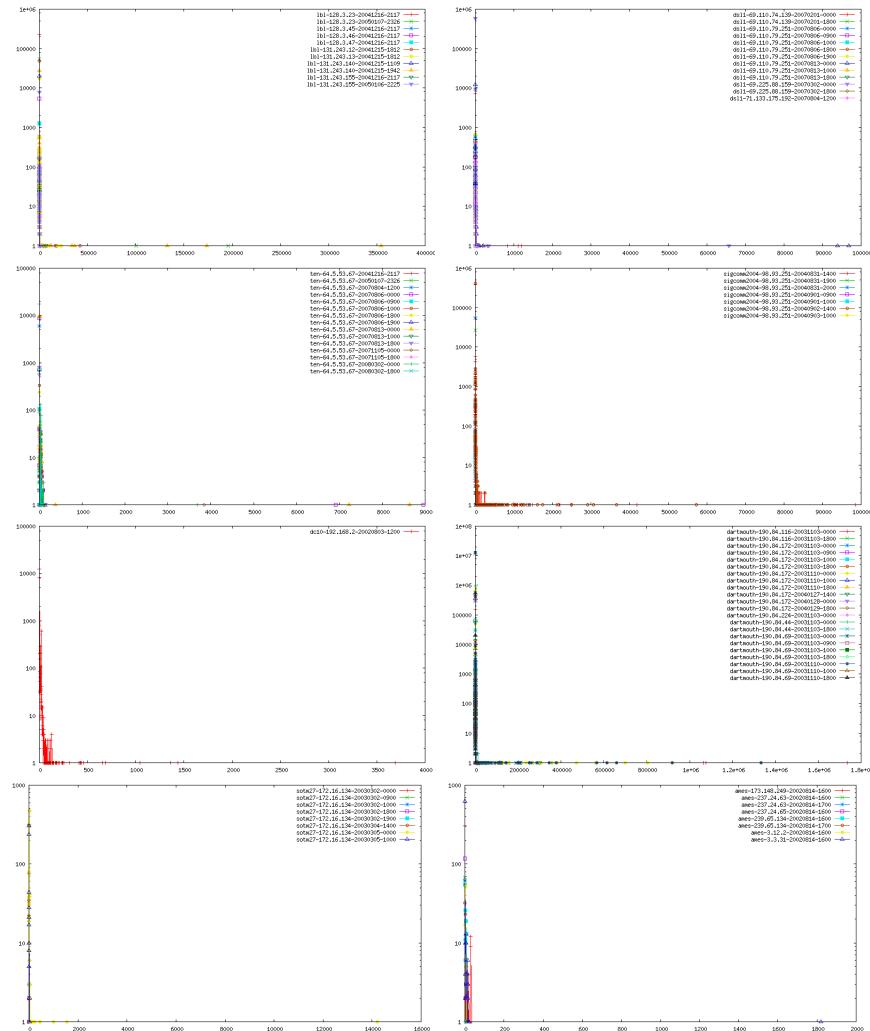


Figure K.59: Number of packets distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

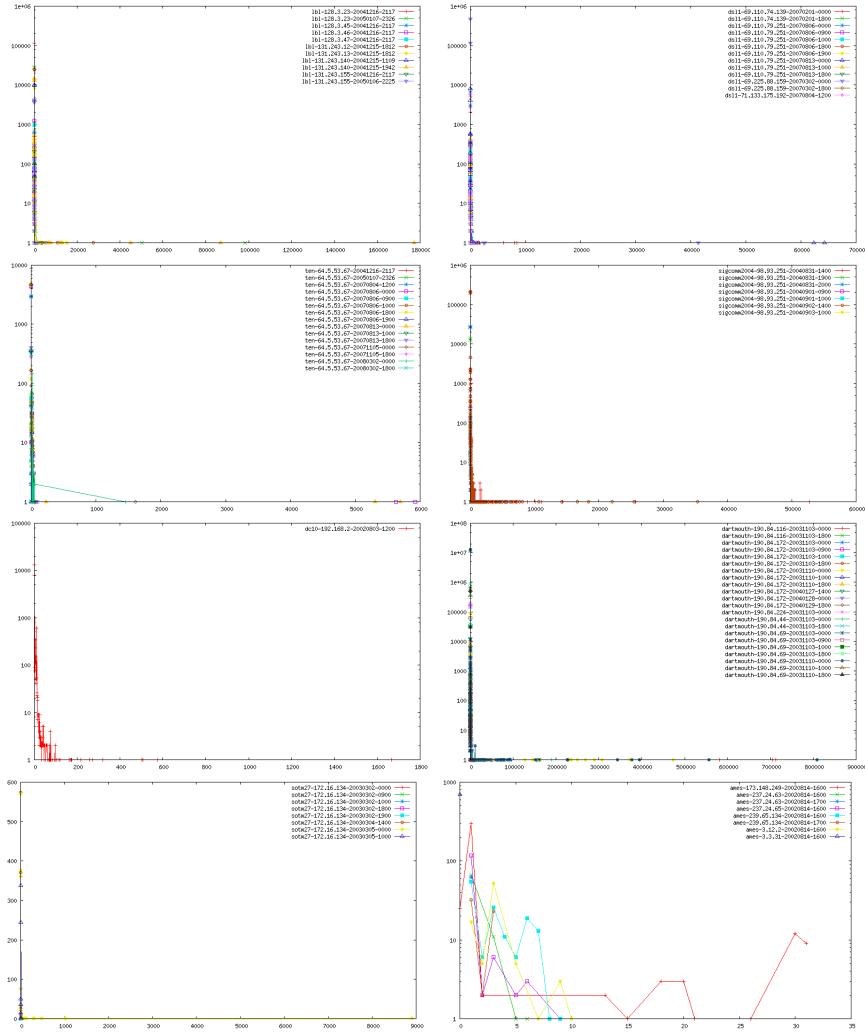


Figure K.60: Number of packets in distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

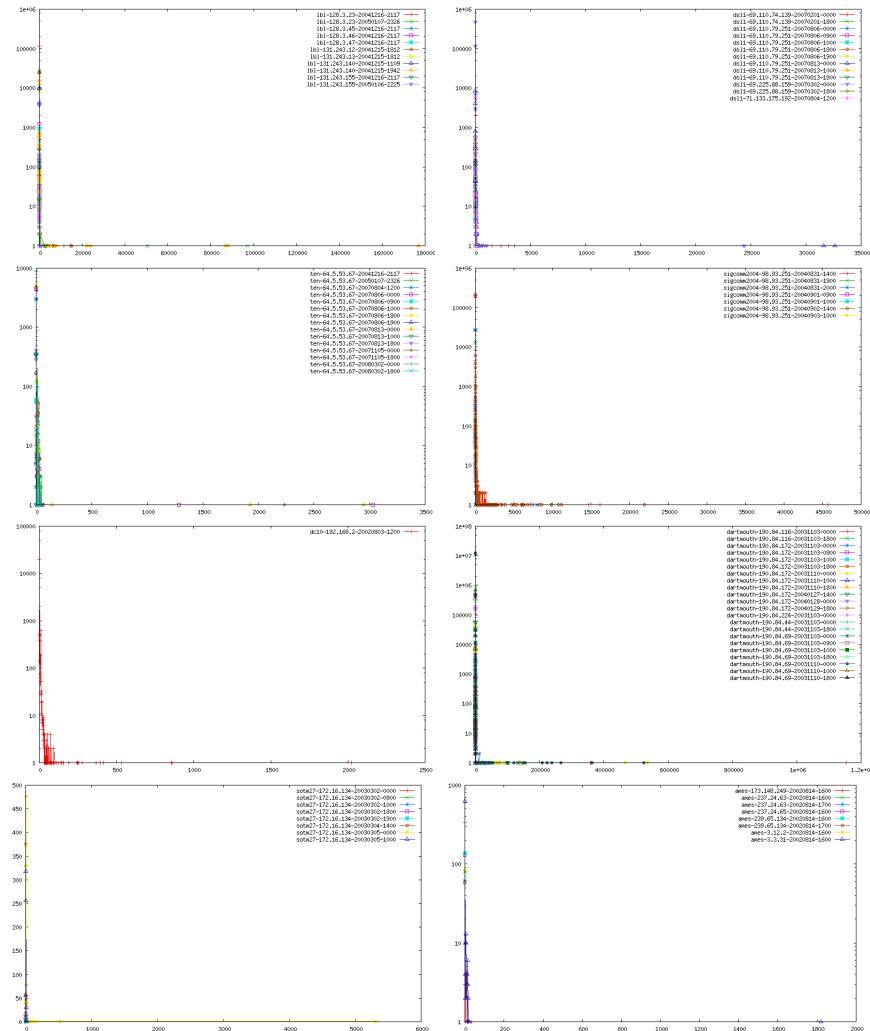


Figure K.61: Number of packets out distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

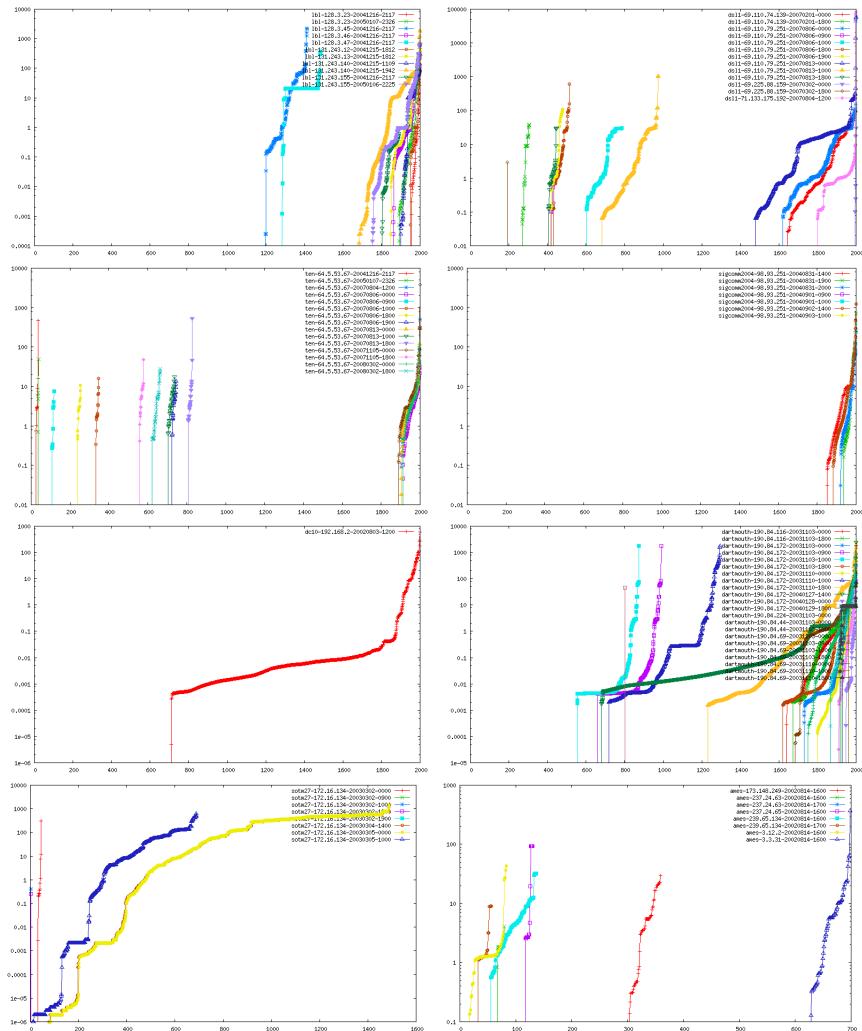


Figure K.62: Duration distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

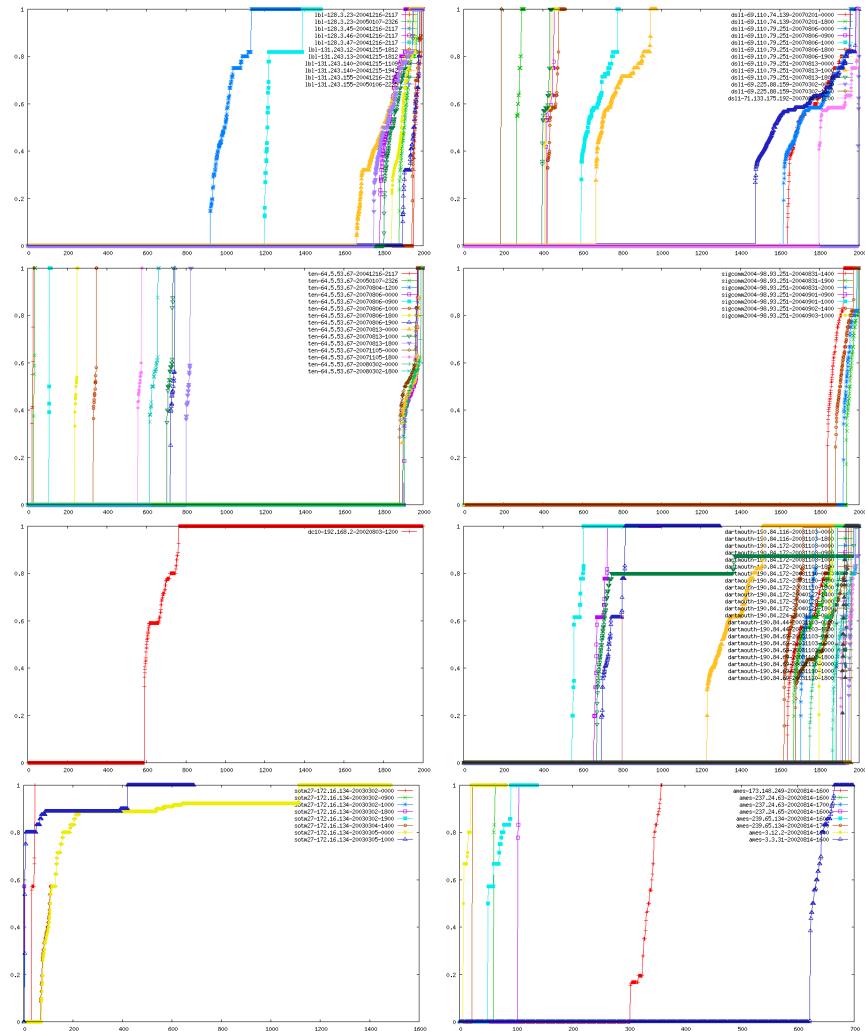


Figure K.63: Number control packets rate distributions for all traces, with all the traces from the same dataset plotted together.

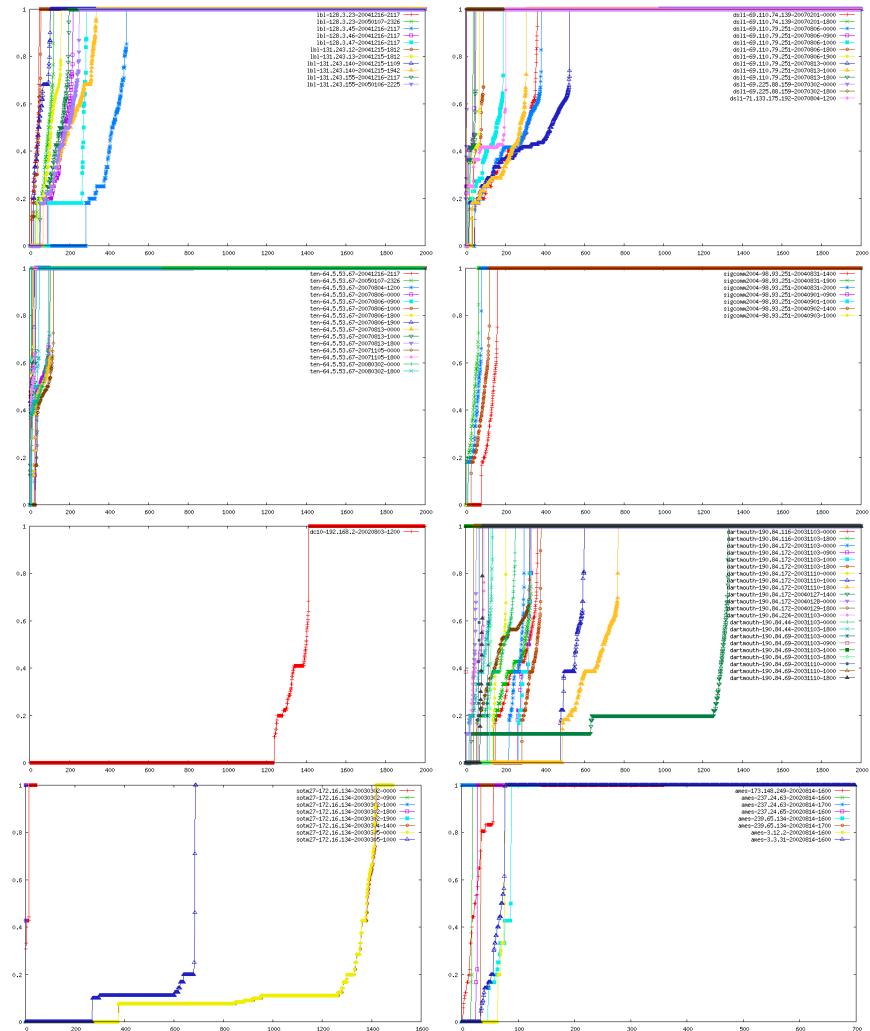


Figure K.64: Number data packets rate distributions for all traces, with all the traces from the same dataset plotted together.

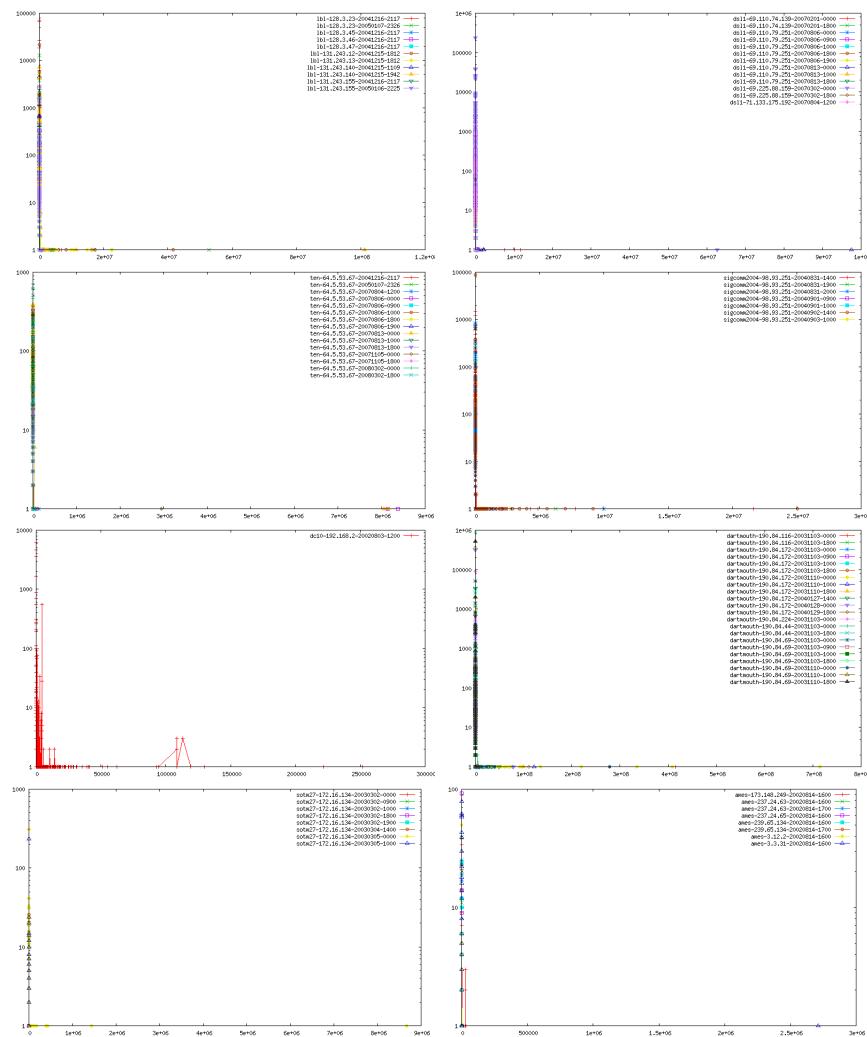


Figure K.65: Number bytes transferred distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

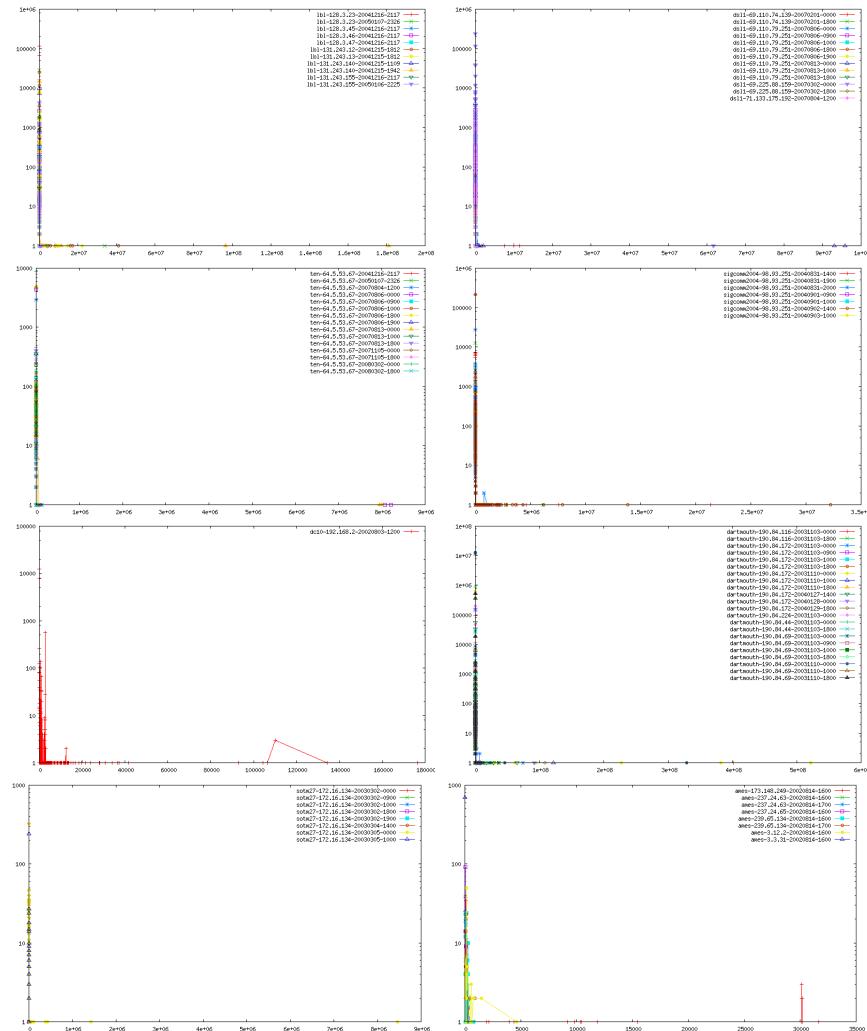


Figure K.66: Number bytes transferred in distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

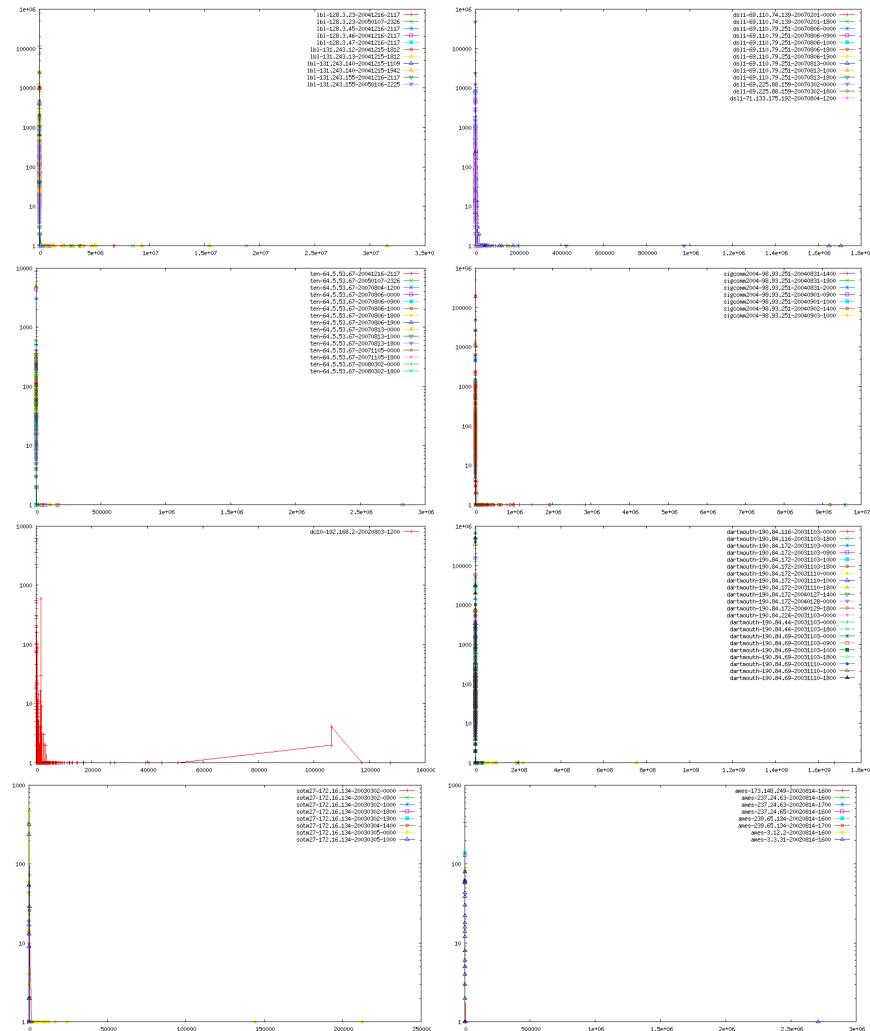


Figure K.67: Number bytes transferred out distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

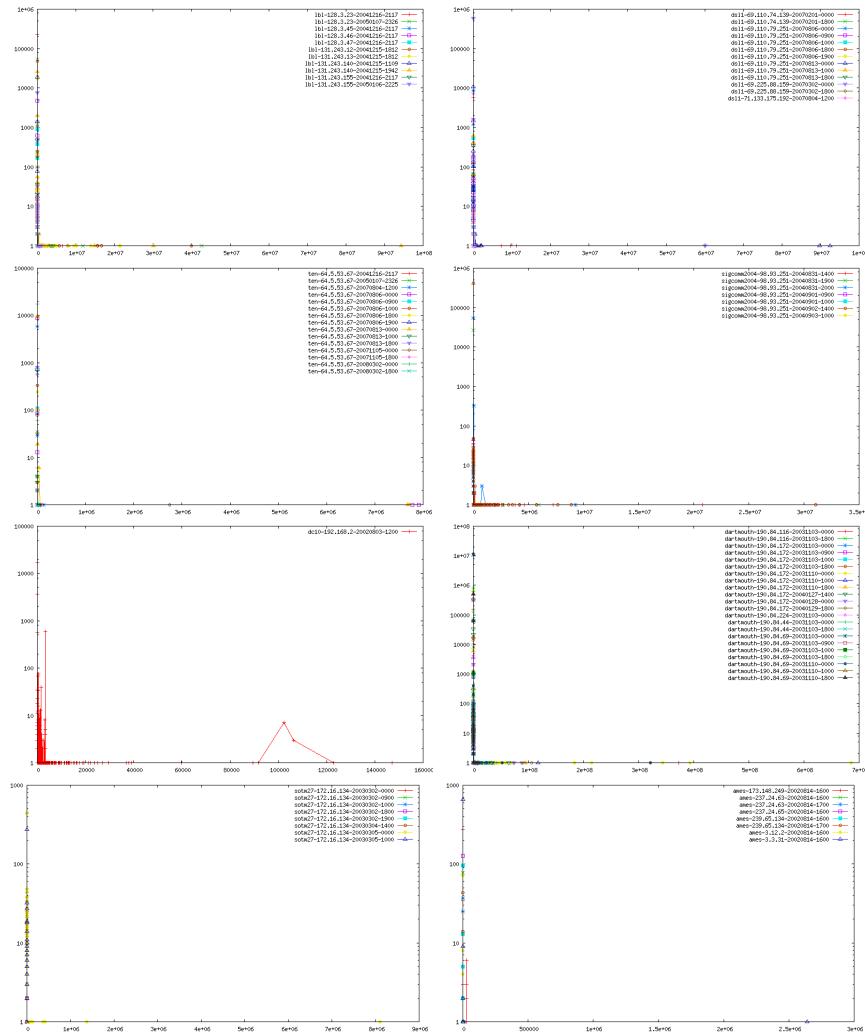


Figure K.68: Number data bytes transferred distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

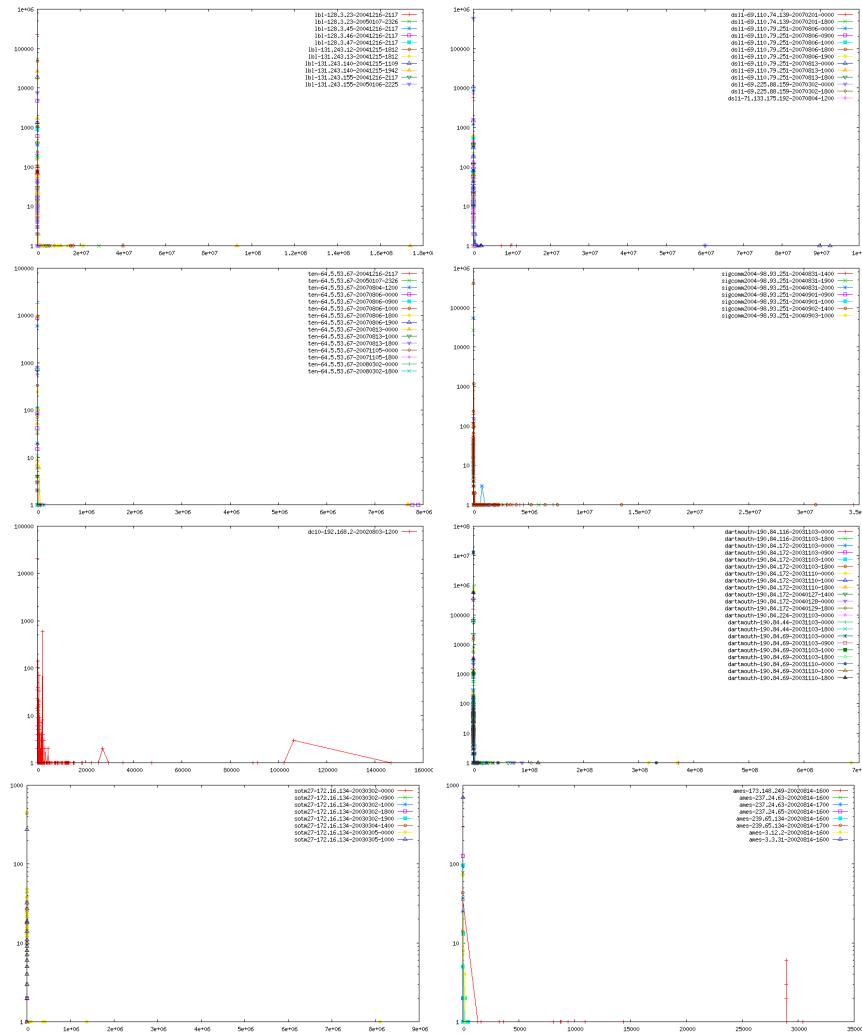


Figure K.69: Number data bytes transferred in distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

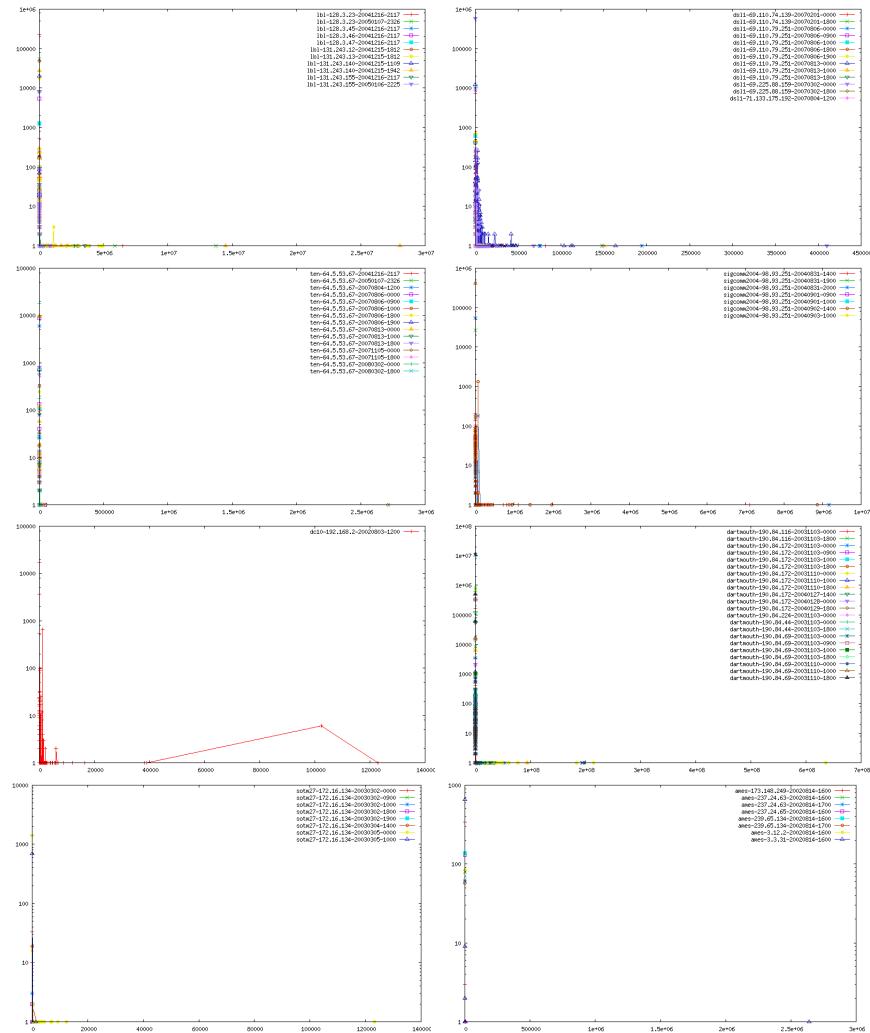


Figure K.70: Number data bytes transferred out distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

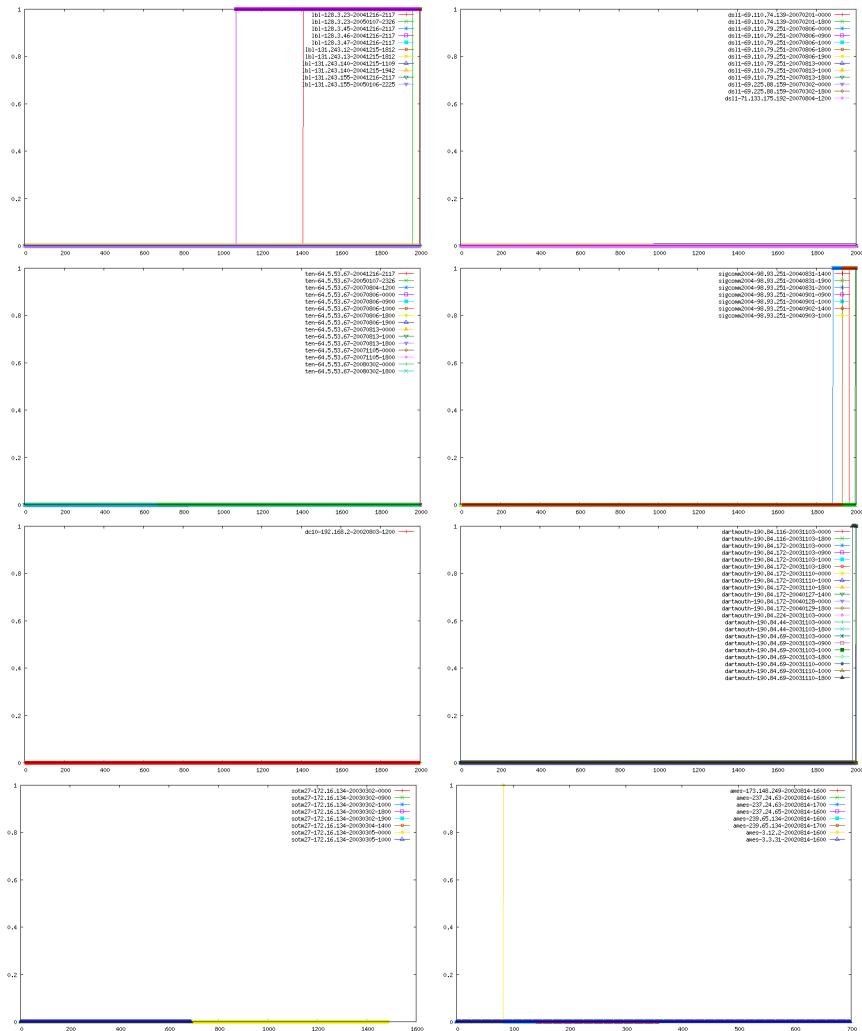


Figure K.71: Fragmented packets rate distributions for all traces, with all the traces from the same dataset plotted together.

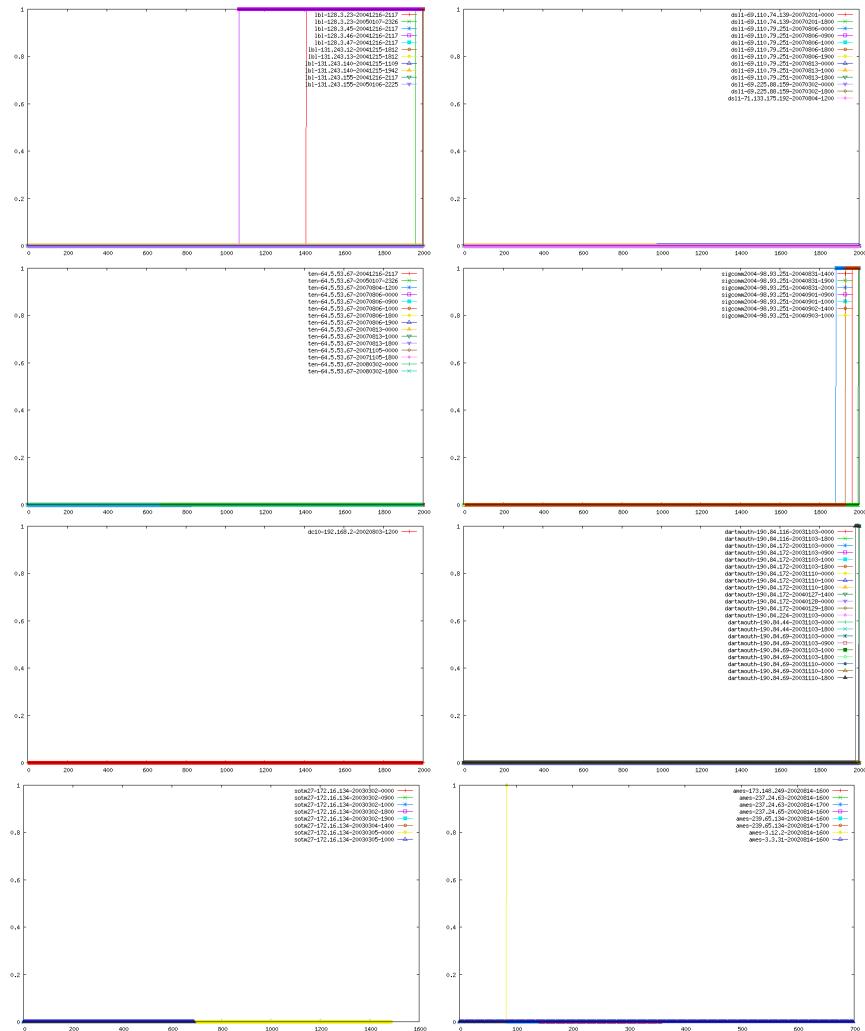


Figure K.72: Bad fragment rate distributions for all traces, with all the traces from the same dataset plotted together.

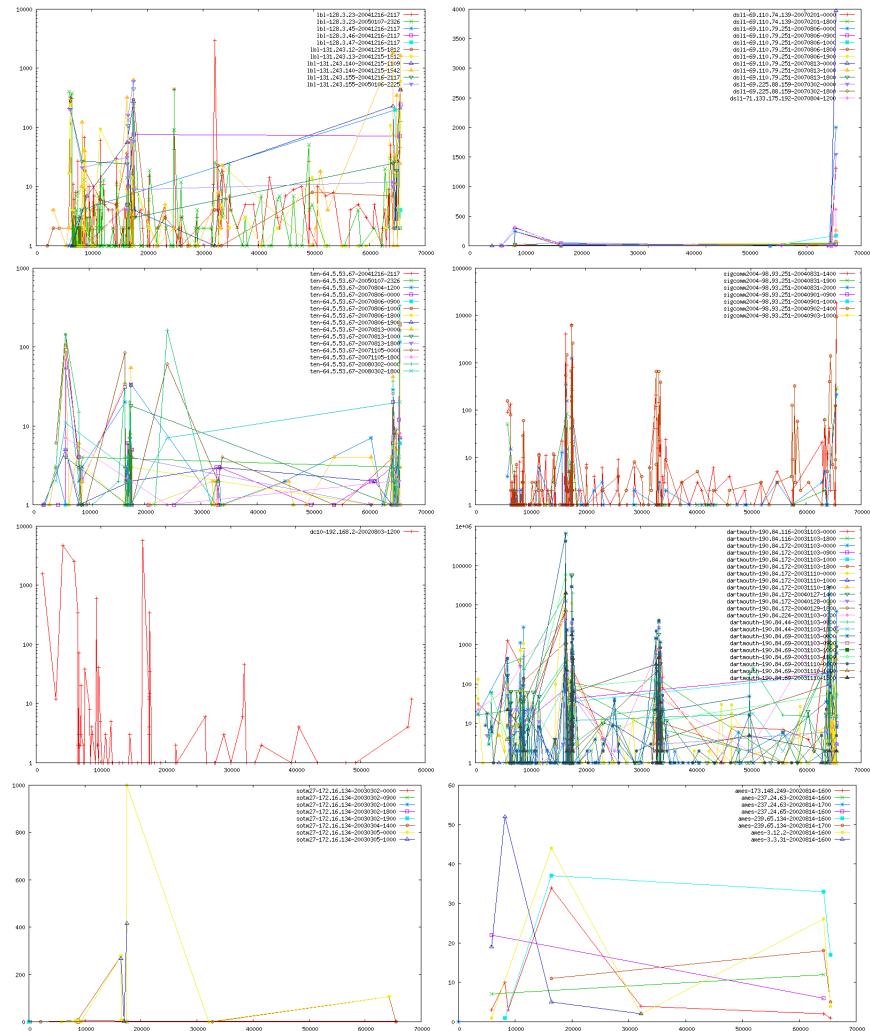


Figure K.73: Max Src Window distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

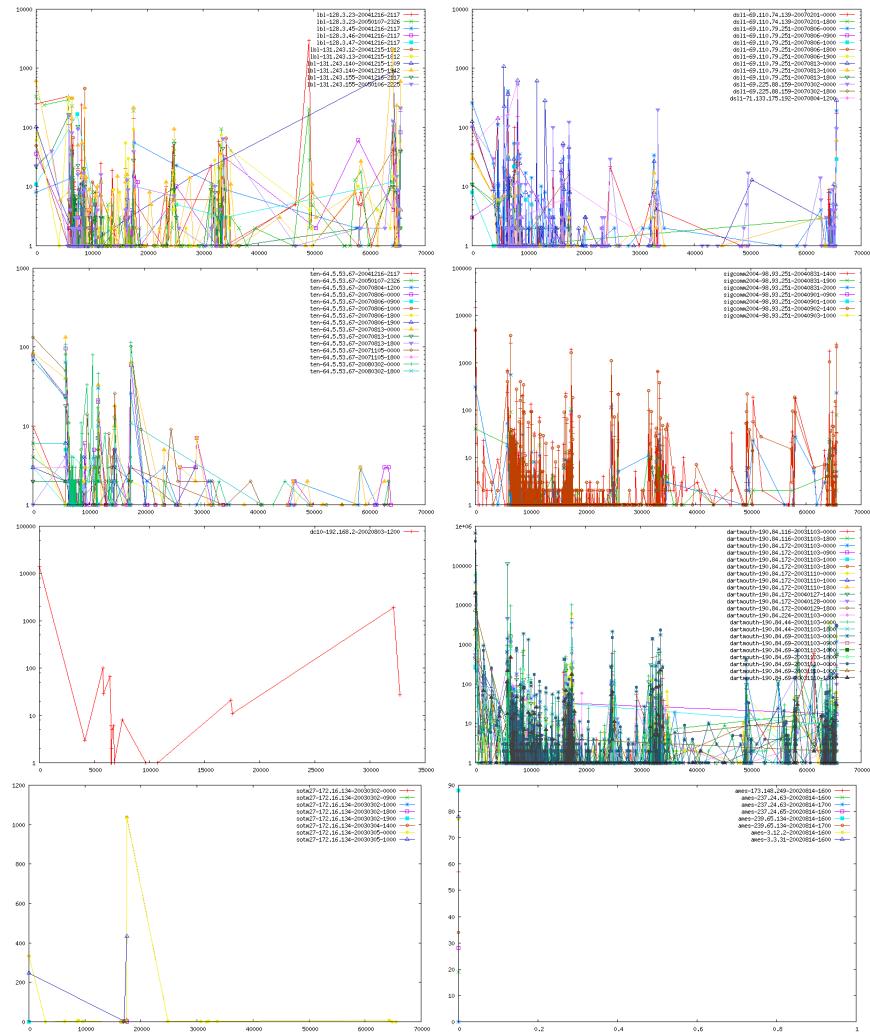


Figure K.74: Max Dst Window distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

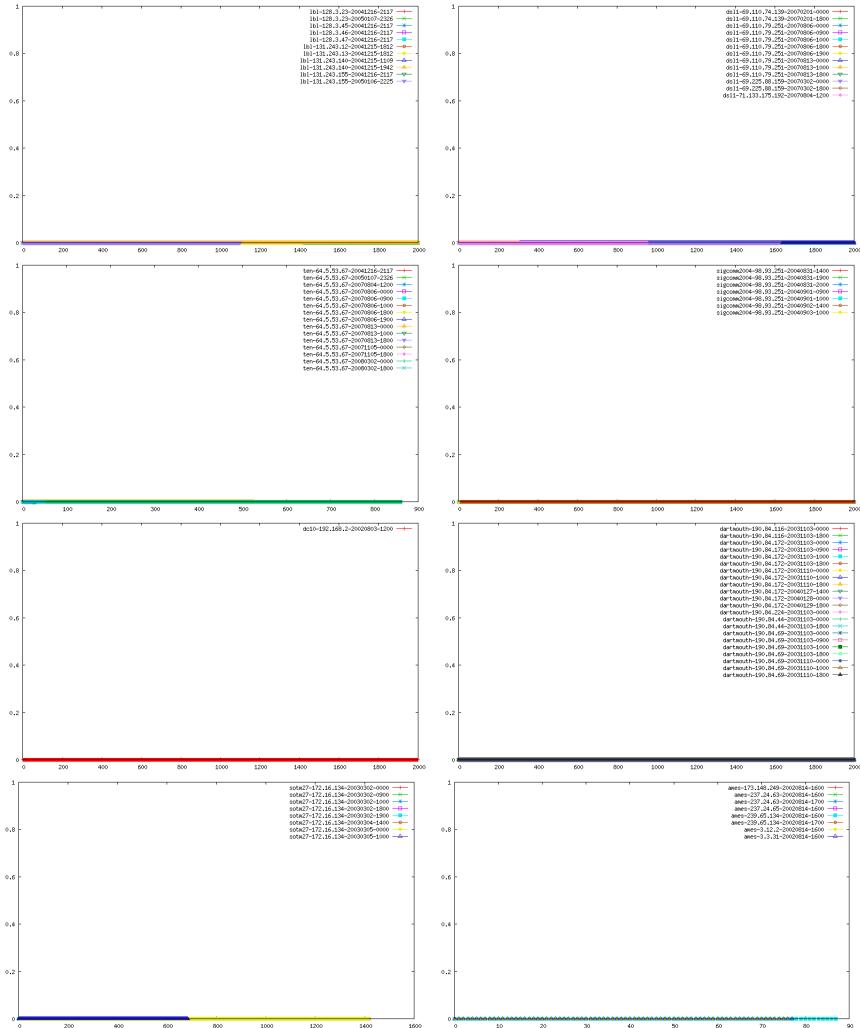


Figure K.75: Urgent rate distributions for all traces, with all the traces from the same dataset plotted together.

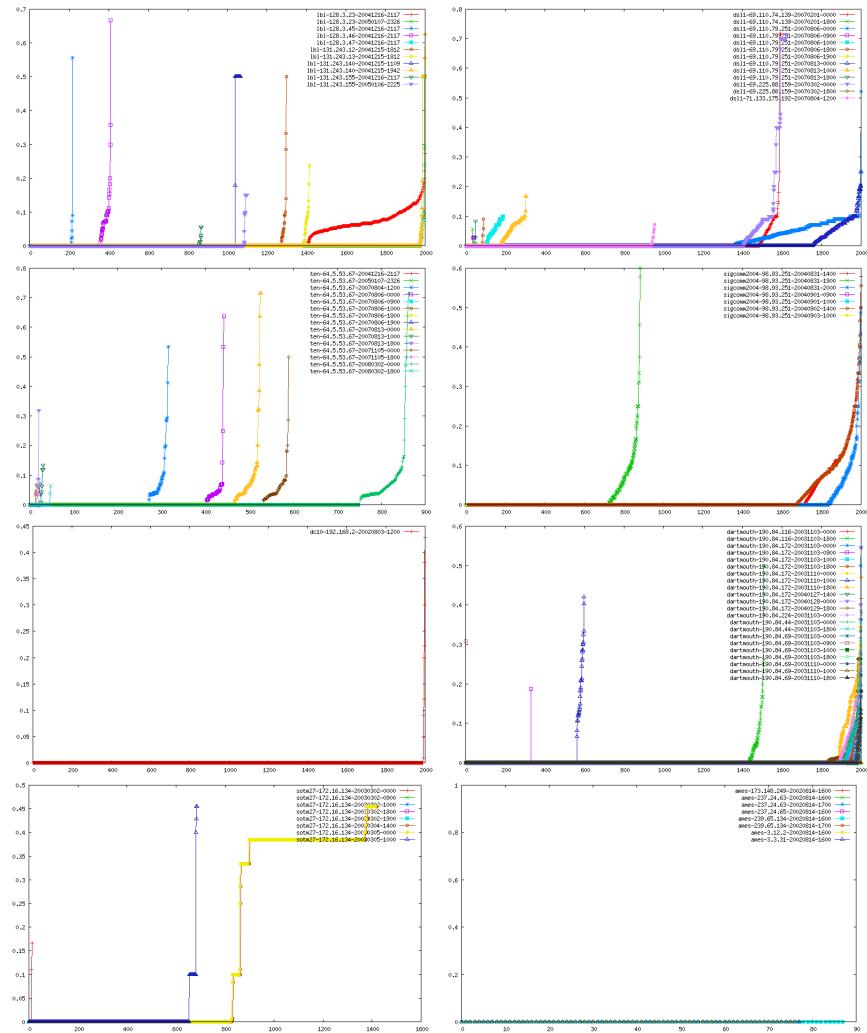


Figure K.76: Resend rate distributions for all traces, with all the traces from the same dataset plotted together.

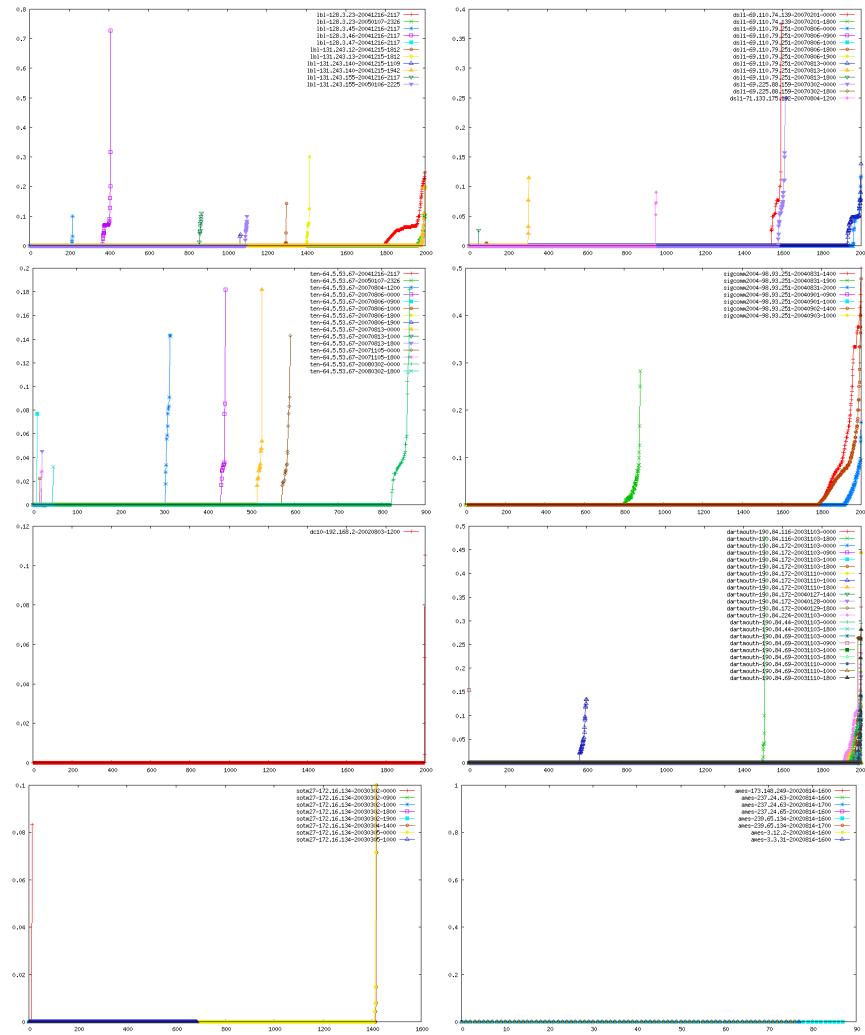


Figure K.77: Wrong resend rate distributions for all traces, with all the traces from the same dataset plotted together.

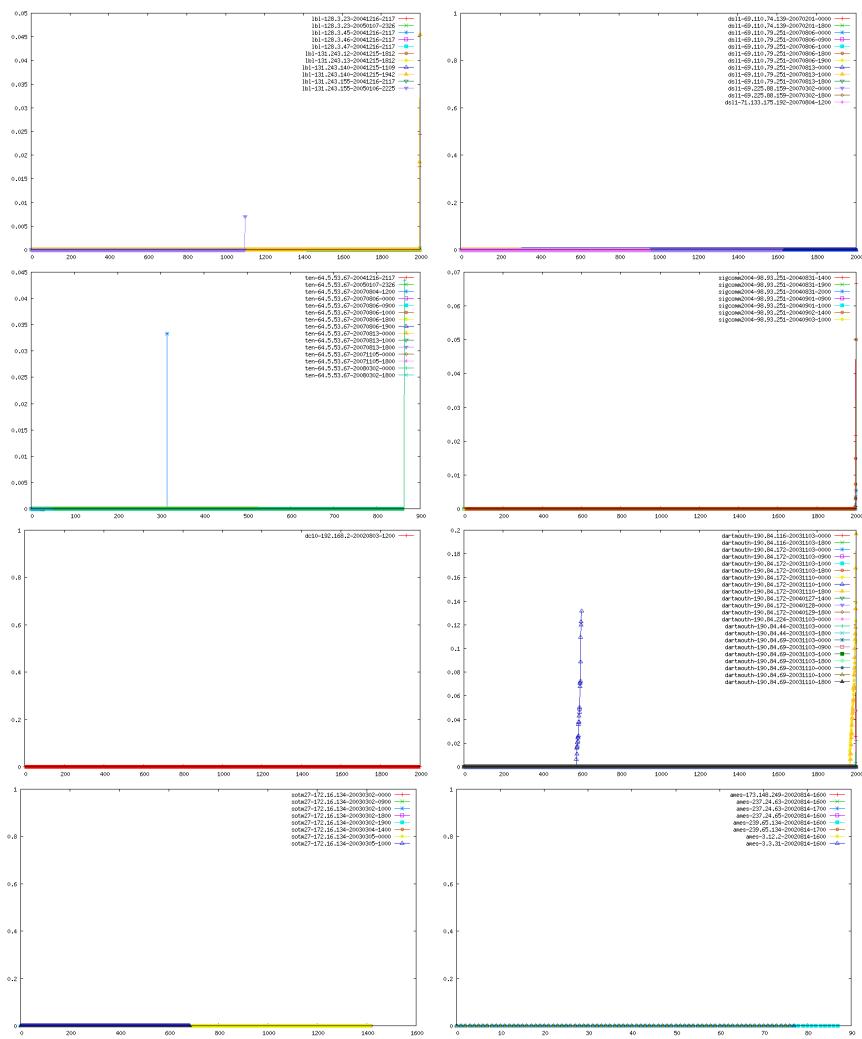


Figure K.78: Duplicate ACK rate distributions for all traces, with all the traces from the same dataset plotted together.

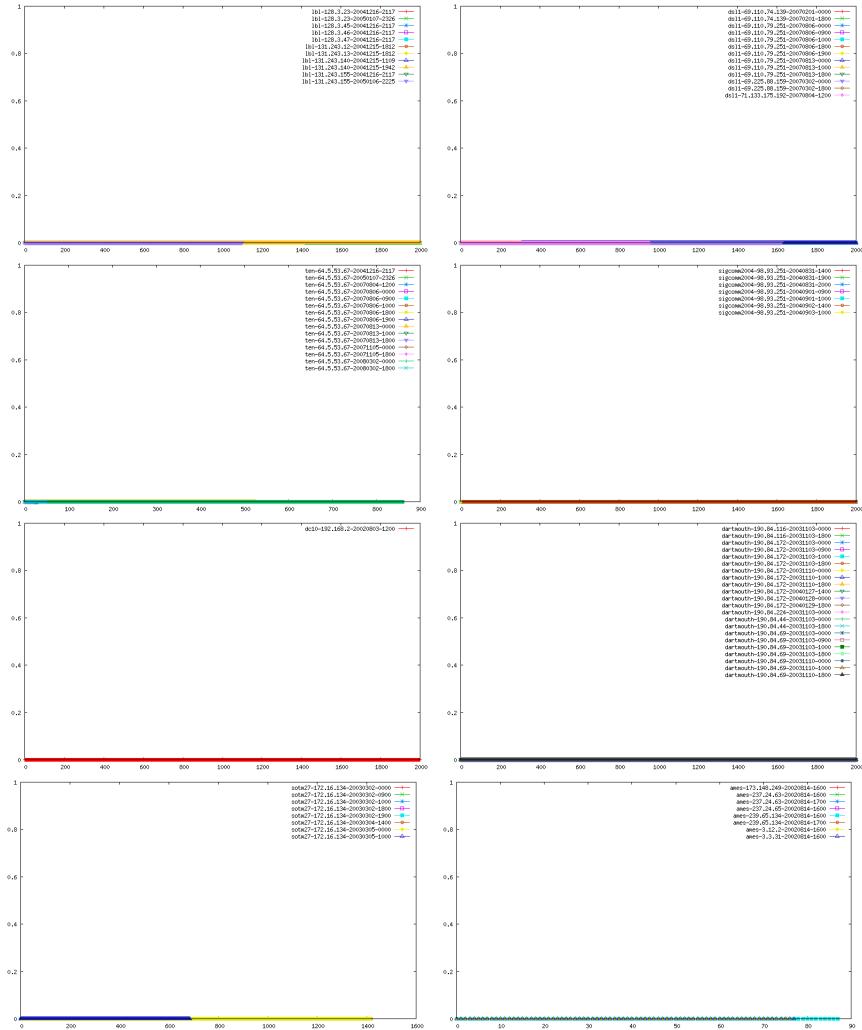


Figure K.79: Wrong ACK distributions for all traces, with all the traces from the same dataset plotted together.

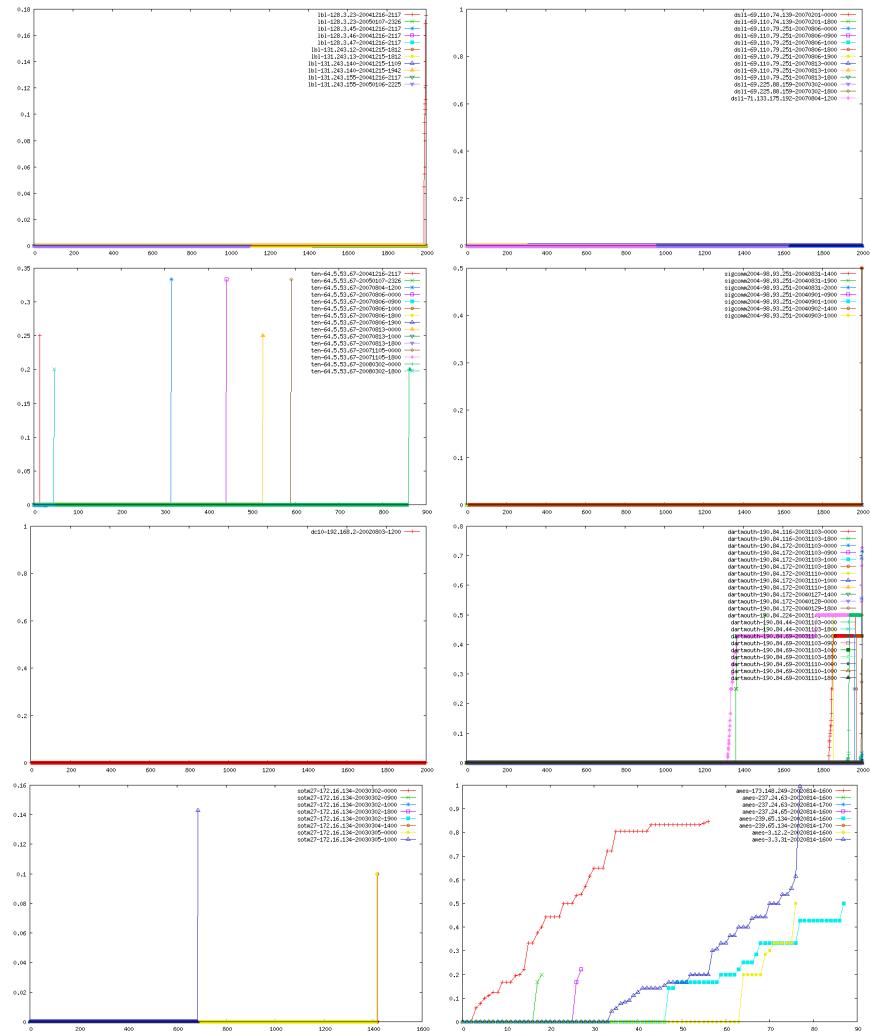


Figure K.80: Wrong data packet size rate distributions for all traces, with all the traces from the same dataset plotted together.

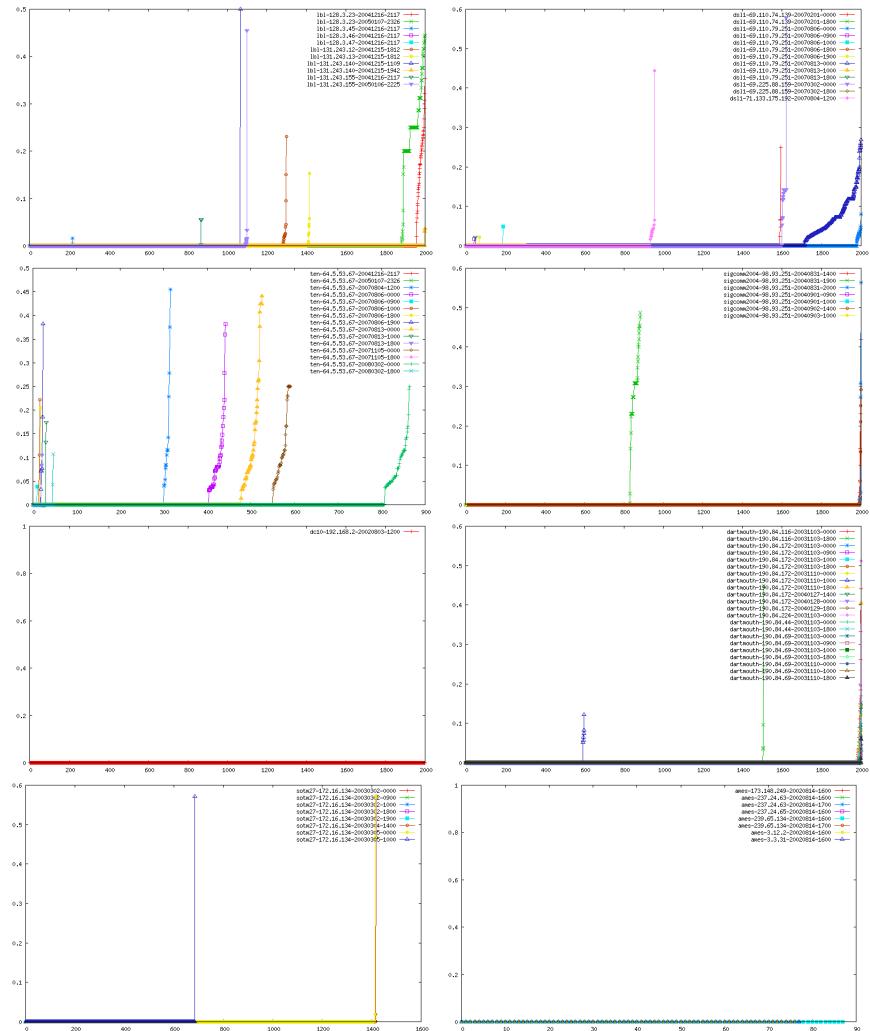


Figure K.81: Window exceeded rate distributions for all traces, with all the traces from the same dataset plotted together.

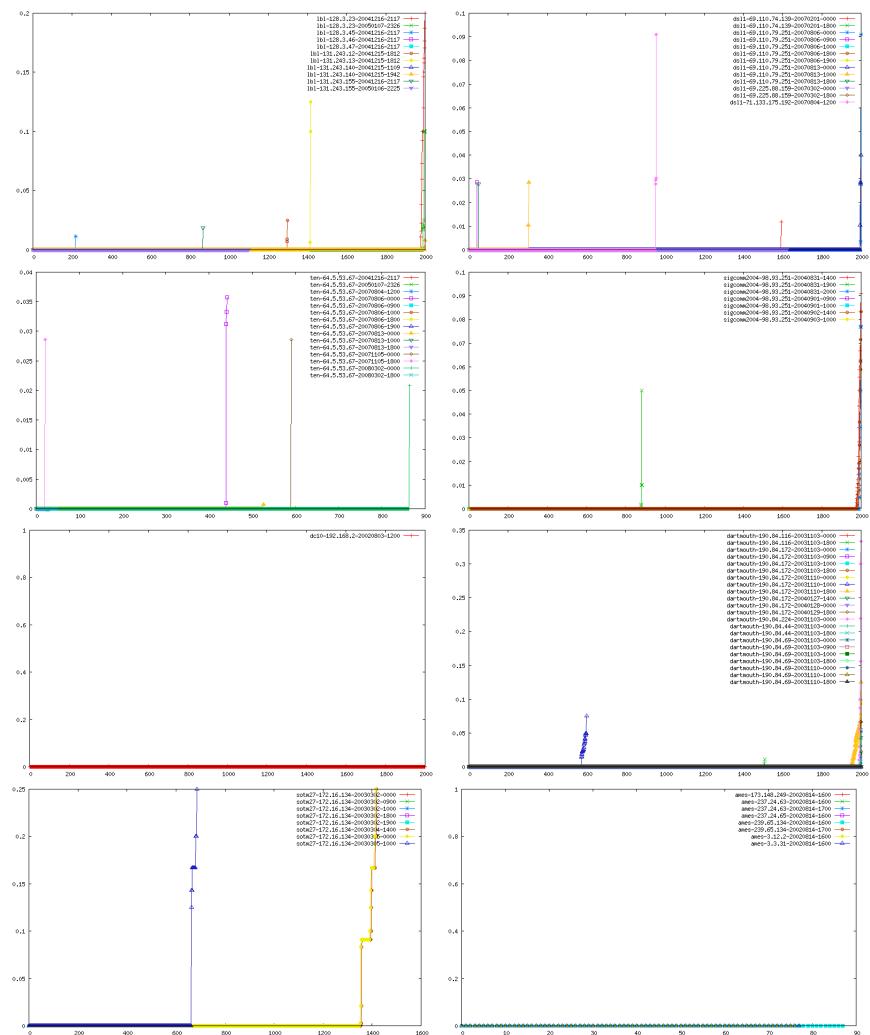


Figure K.82: Hole rate distributions for all traces, with all the traces from the same dataset plotted together.

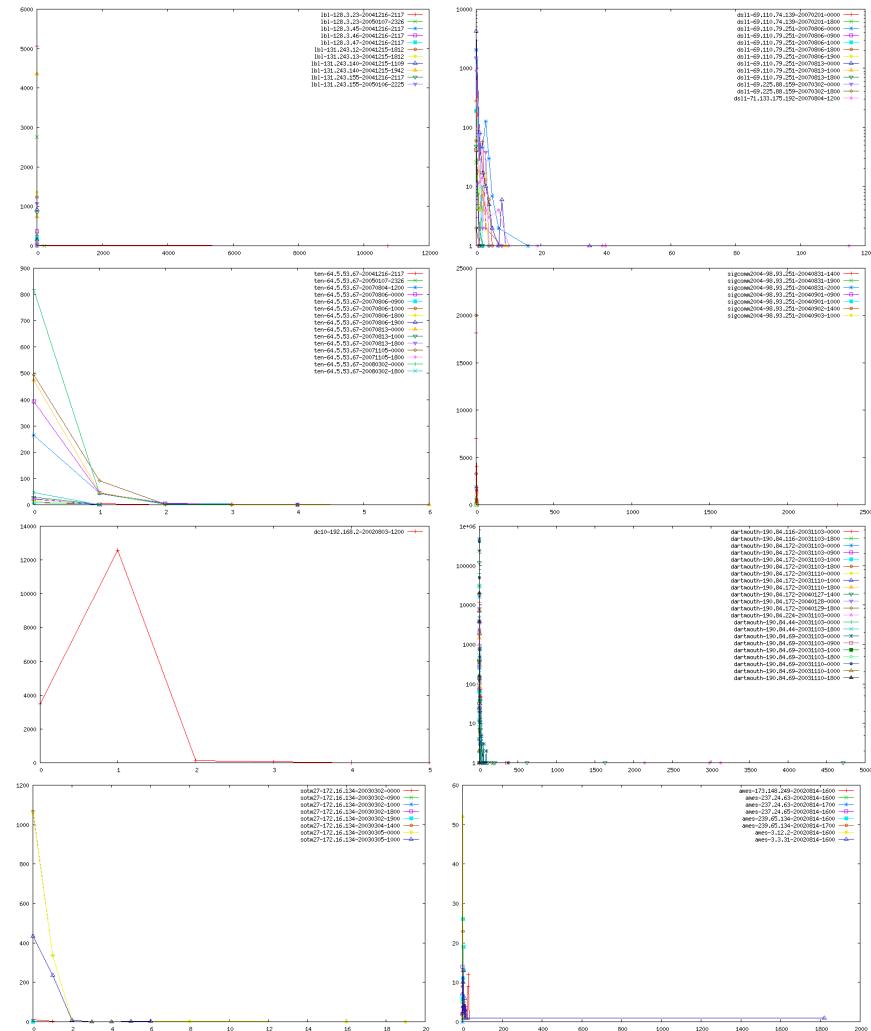


Figure K.83: Number connection errors distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

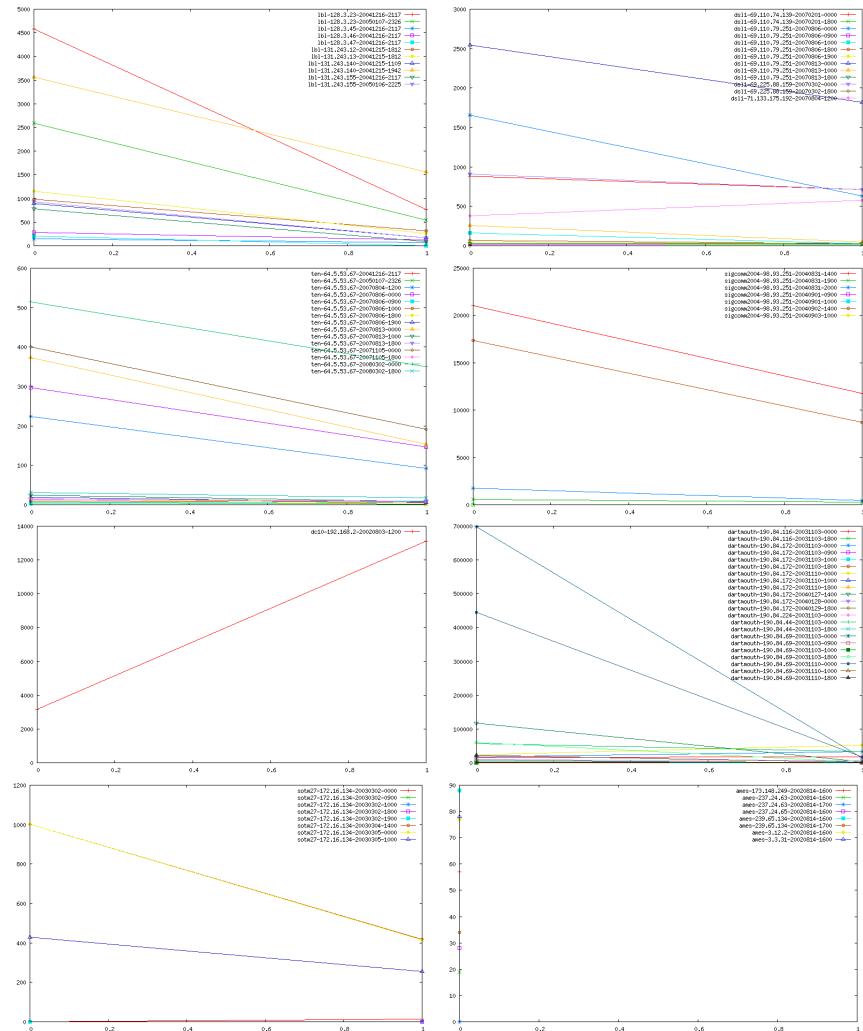


Figure K.84: Number reset connection distributions for all traces, with all the traces from the same dataset plotted together.

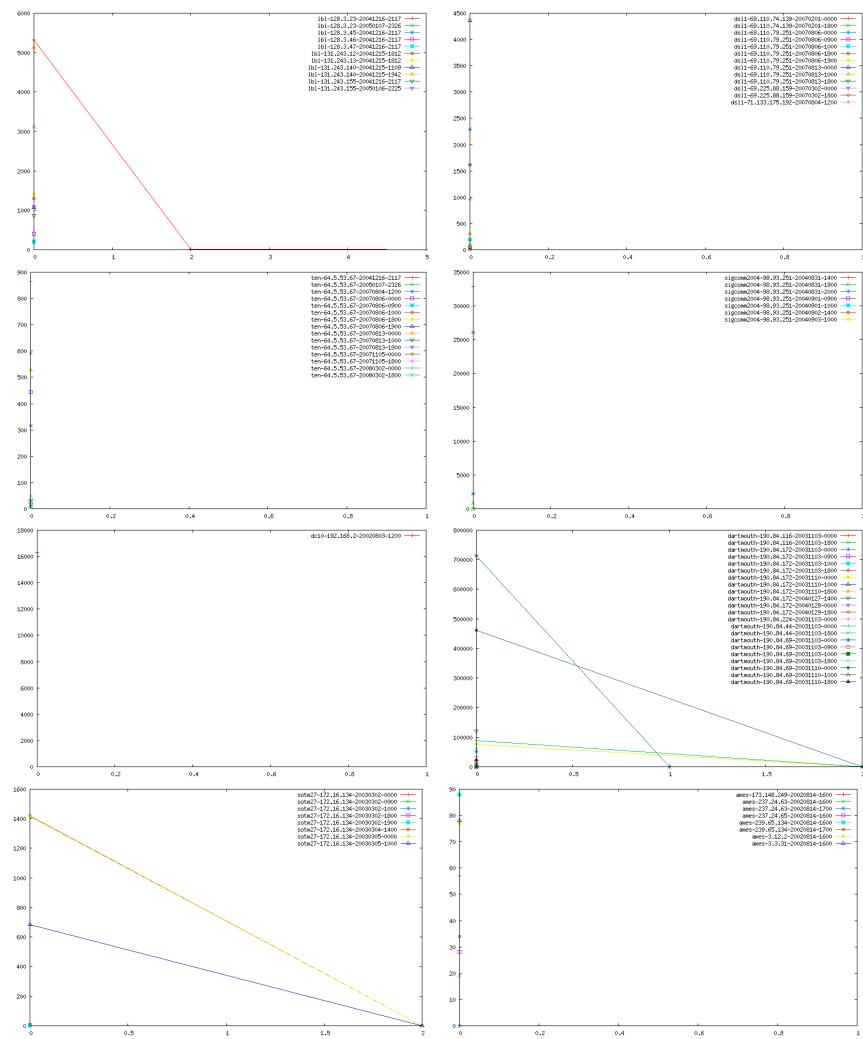


Figure K.85: Number other errors distributions for all traces, with all the traces from the same dataset plotted together.

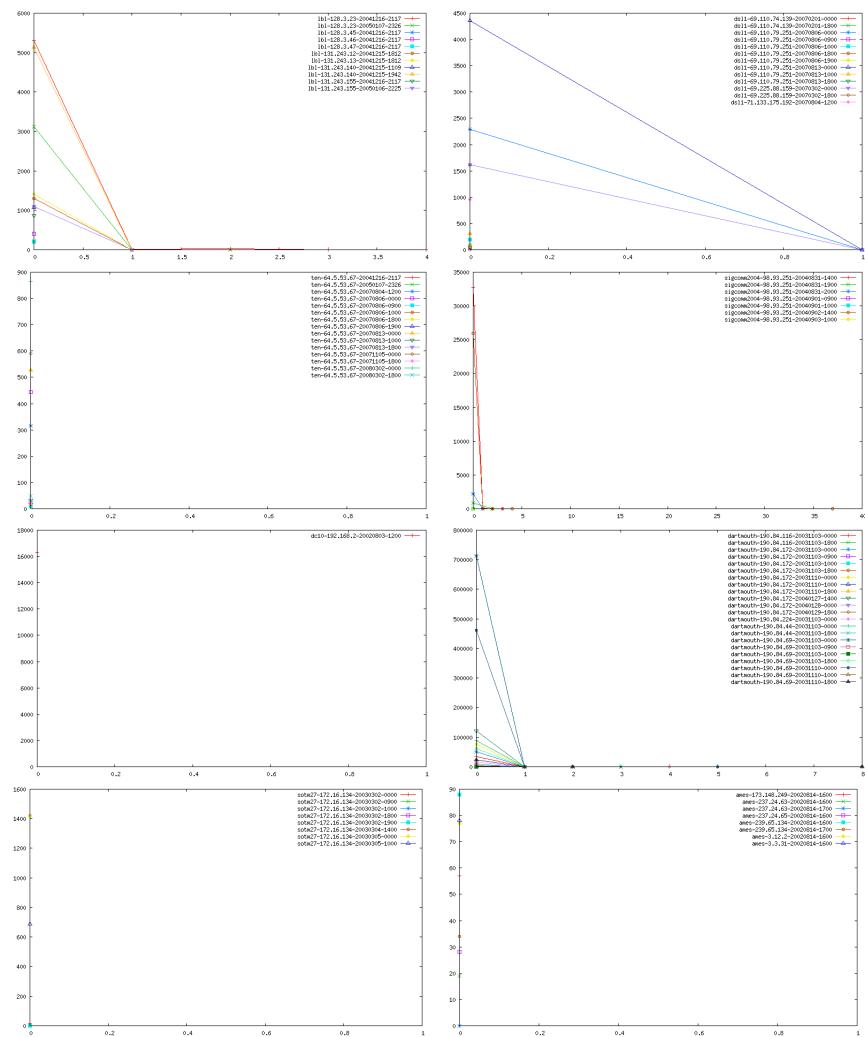


Figure K.86: Number disconnection errors distributions for all traces, with all the traces from the same dataset plotted together.

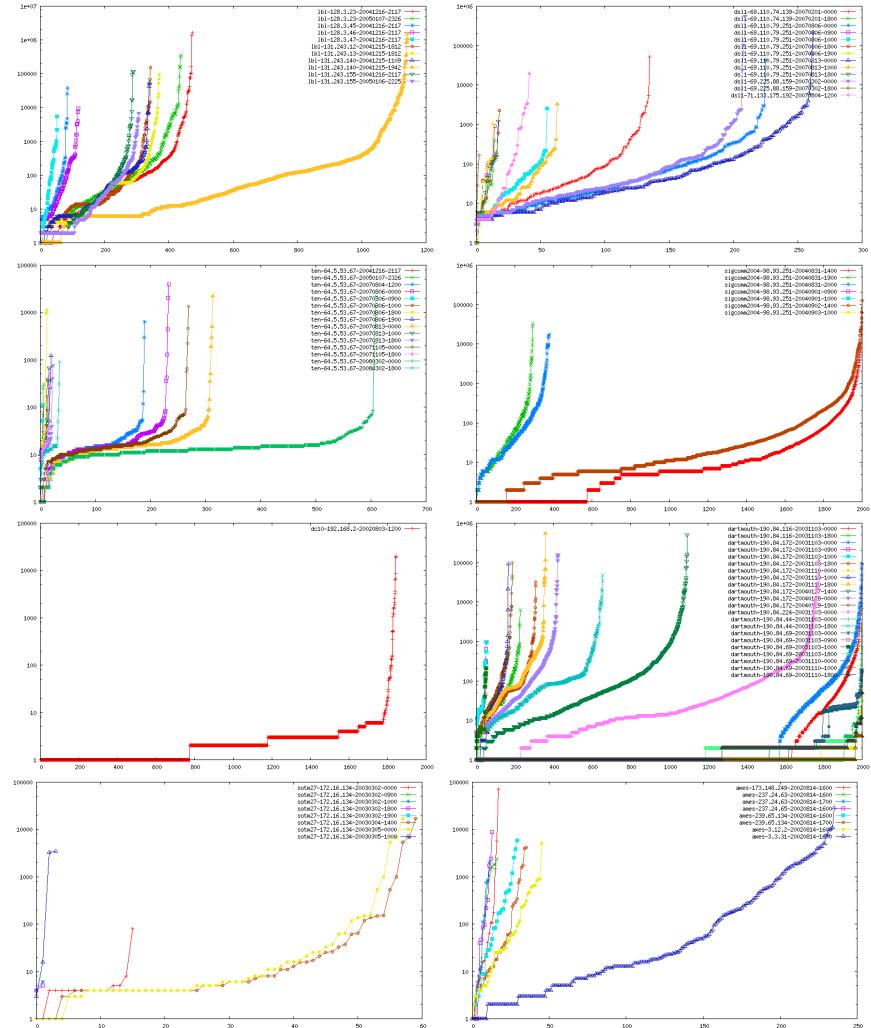


Figure K.87: Packet Destination IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

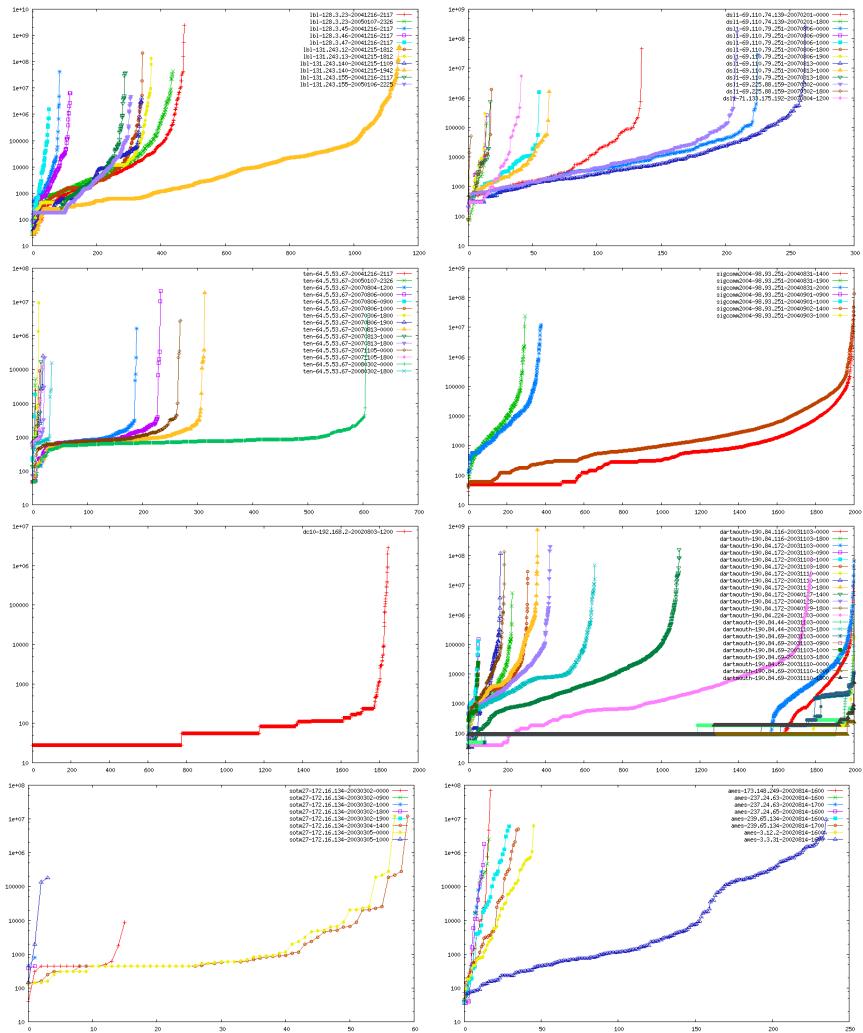


Figure K.88: Bytes Destination IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

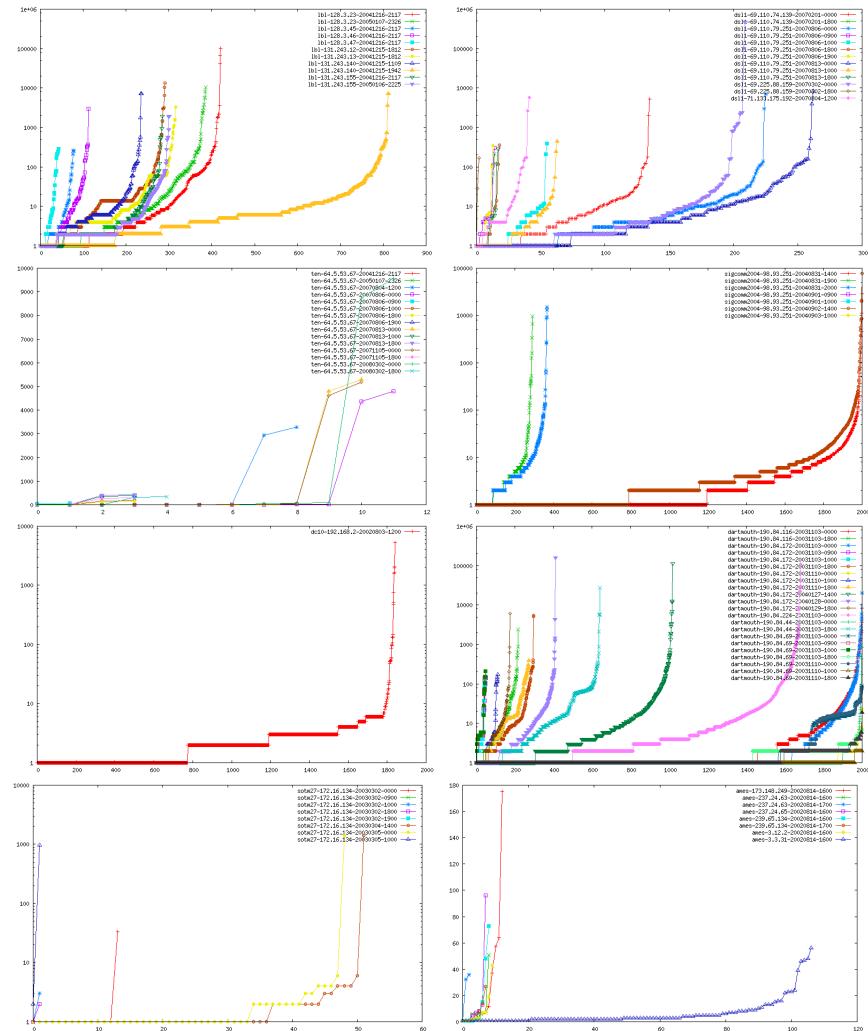


Figure K.89: Connection Destination IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

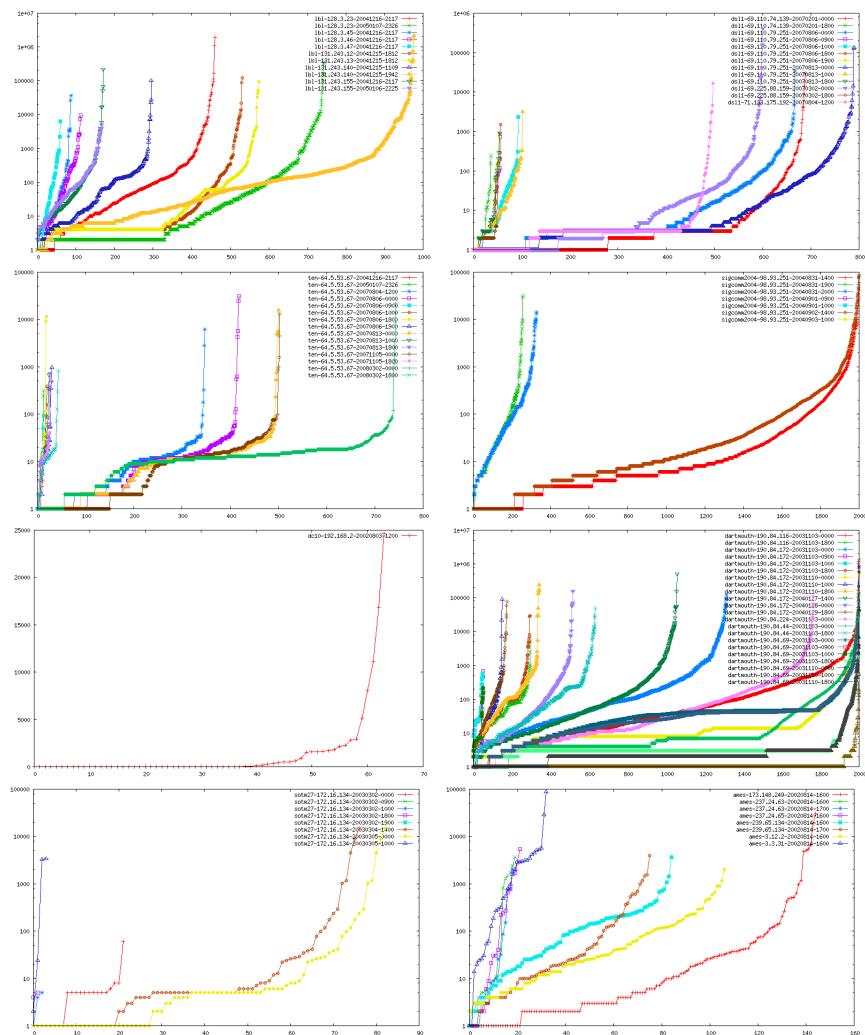


Figure K.90: Packet Source IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

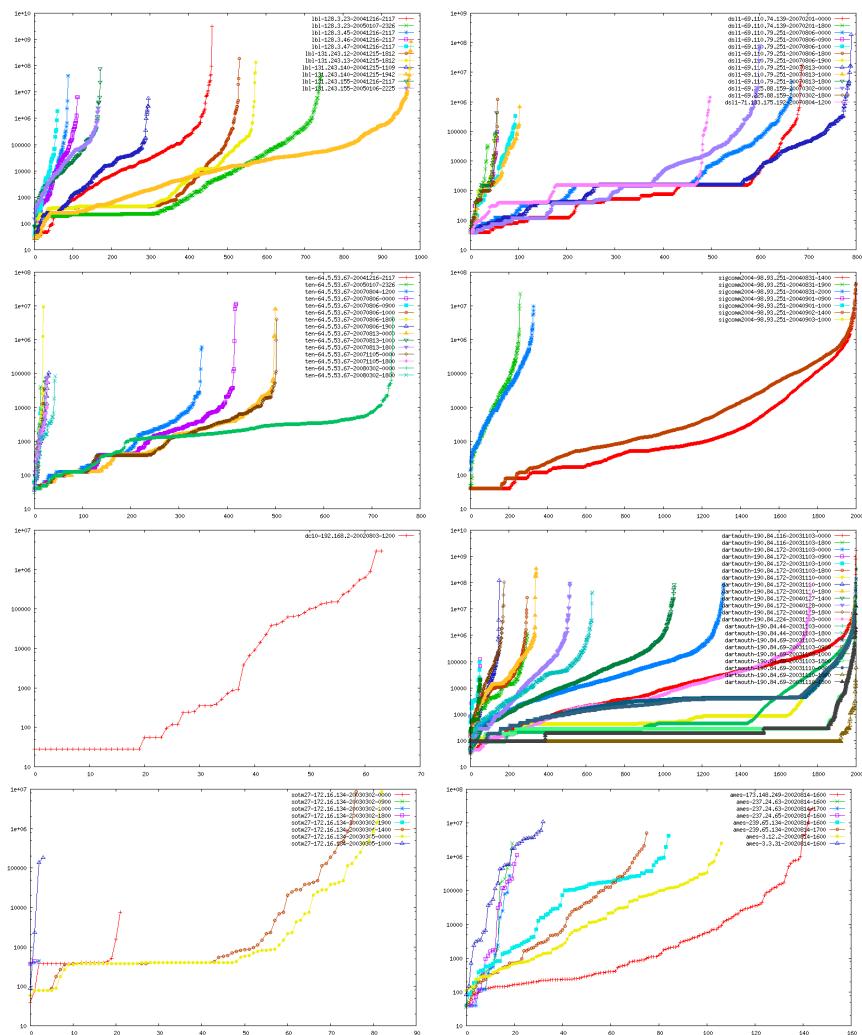


Figure K.91: Bytes Source IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

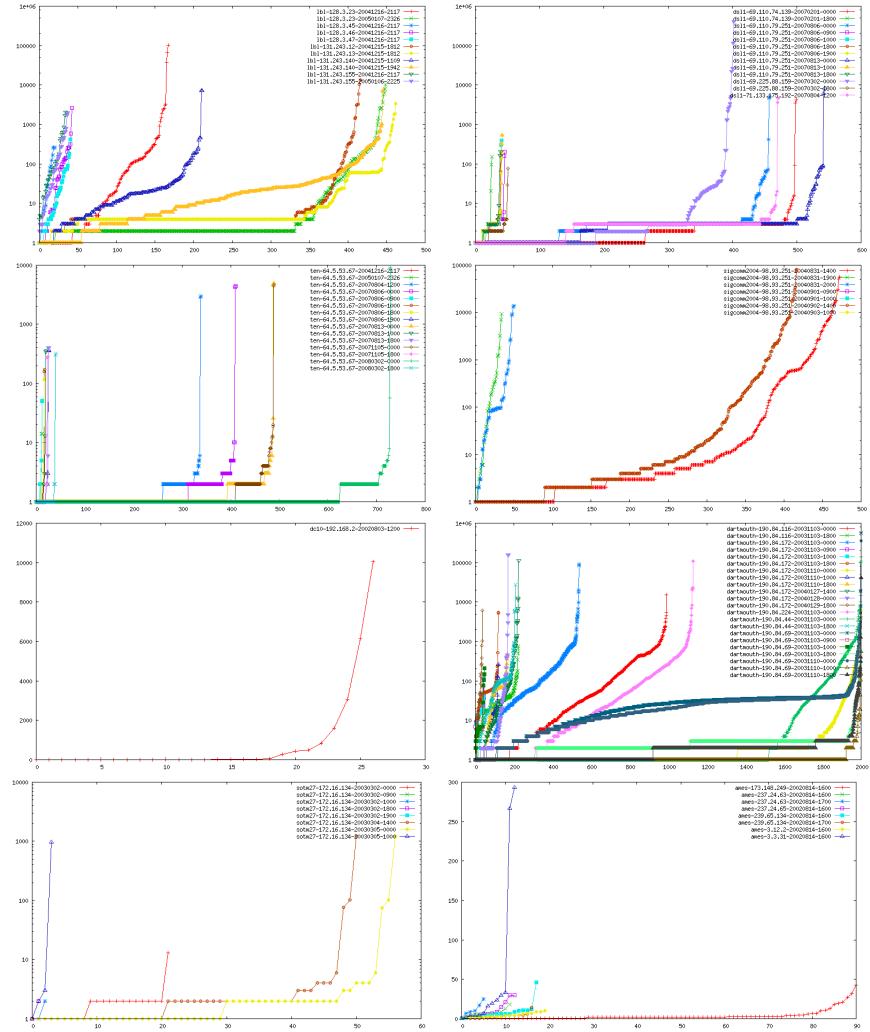


Figure K.92: Connection Source IP distributions for all traces, with all the traces from the same dataset plotted together. Note log scale on some or all of the graphs.

Appendix L

Real data normalized similarities

This appendix presents all the normalized similarities for every metric for the training and testing pairs of all the real data baselines.

Table L.1: Similarity values for individual metrics of baseline 1,
train pair 1

Metric	Similarity
Packets in count	0.227314900723696
Packets out count	0.154428341384863
Connections in count	0.520231213872832
Connections out count	0.496894409937888
Bytes in count	0.32062431280273
Bytes out count	0.0226448646498482
SYN-ONLY rate ratio	0.545454545454546
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.545454545454546
Packet Service discrete	0.00599911070618407
Bytes Service discrete	0.00421385547620304
Connection Service discrete	0.0200070963456337
Packet Source port discrete	0.00401324154204194
Bytes Source port discrete	0.00343595261868428
Continued on next page	

Table L.1 – continued from previous page

Metric	Similarity
Connection Source port discrete	0.00448965887648869
Connection Source port orderedContinuous	0.0162317366715247
Packet TTL discrete	0.17989595185589
Packet TTL orderedContinuous	0.227827291388543
InterPacket delta sortedContinuous	0.5410300727869101
Packet sec orderedContinuous	0.78928304846367
Packet min orderedContinuous	0.218925221287585
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.4166666666666667
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.738323386384151
Bytes min orderedContinuous	0.179419714713109
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.4166666666666667
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.121380495490891
Packets in last w secs orderedContinuous	0.0646592647100104
Priv packets time rate sortedContinuous	0.8939134942204705
Unpriv packets time rate sortedContinuous	0.8820583081533419
Connections time rate sortedContinuous	0.06707378476000628
Priv connections connection time rate sortedContinuous	0.9902930402930402
Unpriv connections connection time rate sortedContinuous	0.9610805860805861
Priv packets priv connection time rate sortedContinuous	0.08616808321248262
Unpriv packets unpriv connection time rate sortedContinuous	0.9610805860805861
SYNs connection time rate sortedContinuous	0.2895012404477042
RSTs connection time rate sortedContinuous	0.9029841061285161
FINs connection time rate sortedContinuous	0.5491736354408205
PSH connection time rate sortedContinuous	0.07816232465000927
Establishment errors connection time rate sortedContinuous	0.9862637362637363
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Continued on next page	

Table L.1 – continued from previous page

Metric	Similarity
Ave duration over last w secs sortedContinuous	0.3461181590214158
Priv packets packet rate sortedContinuous	0.9057635191897476
Unpriv packets packet rate sortedContinuous	0.8960271086175791
InterConnection delta sortedContinuous	0.7662760087167458
Connection sec orderedContinuous	0.363305267351783
Connection min orderedContinuous	0.228736922723112
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.4166666666666667
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.3767413481319463
Connection Priv connections rate sortedContinuous	0.9948495111185859
Connection Unpriv connections rate sortedContinuous	0.8
Connection Priv packet rate sortedContinuous	0.3768262462728213
Connection Unpriv packet rate sortedContinuous	0.8
Connection SYNs rate sortedContinuous	0.5342196668957315
Connection RSTs rate sortedContinuous	0.8706449683836141
Connection FINs rate sortedContinuous	0.7857209768914176
Connection PSH rate sortedContinuous	0.3031840613433473
Connection Establishment errors rate sortedContinuous	0.8026143790849673
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.6797967970728512
Number of packets orderedContinuous	0.441109512723567
Number of packets in orderedContinuous	0.510683331522553
Number of packets out orderedContinuous	0.532073411683061
Duration sortedContinuous	0.959072946333819
Number control packets rate sortedContinuous	0.9556241731724588
Number data packets rate sortedContinuous	0.9684543513994418
Number bytes transferred orderedContinuous	0.283663799729125
Number bytes transferred in orderedContinuous	0.247300949670787
Number bytes transferred out orderedContinuous	0.186387895637763
Continued on next page	

Table L.1 – continued from previous page

Metric	Similarity
Number data bytes transferred orderedContinuous	0.0659142157894491
Number data bytes transferred in orderedContinuous	0.0656599308828495
Number data bytes transferred out orderedContinuous	0.12343900571514
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.213145539906103
Max Dst Window orderedContinuous	0.132863021751911
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.8827160493827161
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.7610537463595446
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.588205517282289
Number reset connection orderedContinuous	0.632417522457728
Number other errors orderedContinuous	0.62557497700092
Number disconnection errors orderedContinuous	0.62557497700092
Packet Destination IP nonKeyedSortedContinuous	0.710103929728151
Bytes Destination IP nonKeyedSortedContinuous	0.69935571328369
Connection Destination IP nonKeyedSortedContinuous	0.668573209630673
Packet Source IP nonKeyedSortedContinuous	0.75370447347218
Bytes Source IP nonKeyedSortedContinuous	0.699059541447835
Connection Source IP nonKeyedSortedContinuous	0.930400374026772

Table L.2: Similarity values for individual metrics of basemode 1,
train pair 2

Metric	Similarity
Packets in count	0
Packets out count	0
Connections in count	0
Connections out count	0
Bytes in count	0
Bytes out count	0
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	0
Bytes Service discrete	0
Connection Service discrete	0
Packet Source port discrete	0
Bytes Source port discrete	0
Connection Source port discrete	0
Connection Source port orderedContinuous	0
Packet TTL discrete	0
Packet TTL orderedContinuous	0
InterPacket delta sortedContinuous	0
Packet sec orderedContinuous	0
Packet min orderedContinuous	0
Packet GmHour orderedContinuous	0
Packet LocHour orderedContinuous	0
Packet weekday orderedContinuous	0
Bytes sec orderedContinuous	0
Bytes min orderedContinuous	0
Bytes GmHour orderedContinuous	0
Bytes LocHour orderedContinuous	0
Continued on next page	

Table L.2 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	0
Packet size orderedContinuous	0
Packets in last w secs orderedContinuous	0
Priv packets time rate sortedContinuous	0
Unpriv packets time rate sortedContinuous	0
Connections time rate sortedContinuous	0
Priv connections connection time rate sortedContinuous	0
Unpriv connections connection time rate sortedContinuous	0
Priv packets priv connection time rate sortedContinuous	0
Unpriv packets unpriv connection time rate sortedContinuous	0
SYNs connection time rate sortedContinuous	0
RSTs connection time rate sortedContinuous	0
FINs connection time rate sortedContinuous	0
PSH connection time rate sortedContinuous	0
Establishment errors connection time rate sortedContinuous	0
Other errors connection time rate sortedContinuous	0
Disconnection errors connection time rate sortedContinuous	0
Ave duration over last w secs sortedContinuous	0
Priv packets packet rate sortedContinuous	0
Unpriv packets packet rate sortedContinuous	0
InterConnection delta sortedContinuous	0
Connection sec orderedContinuous	0
Connection min orderedContinuous	0
Connection GmHour orderedContinuous	0
Connection LocHour orderedContinuous	0
Connection weekday orderedContinuous	0
Connection packet rate sortedContinuous	0
Connection Priv connections rate sortedContinuous	0
Connection Unpriv connections rate sortedContinuous	0
Connection Priv packet rate sortedContinuous	0
Connection Unpriv packet rate sortedContinuous	0
Continued on next page	

Table L.2 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0
Connection RSTs rate sortedContinuous	0
Connection FINs rate sortedContinuous	0
Connection PSH rate sortedContinuous	0
Connection Establishment errors rate sortedContinuous	0
Connection Other errors rate sortedContinuous	0
Connection Disconnection errors rate sortedContinuous	0
Ave duration over last m connections sortedContinuous	0
Number of packets orderedContinuous	0
Number of packets in orderedContinuous	0
Number of packets out orderedContinuous	0
Duration sortedContinuous	0
Number control packets rate sortedContinuous	0
Number data packets rate sortedContinuous	0
Number bytes transferred orderedContinuous	0
Number bytes transferred in orderedContinuous	0
Number bytes transferred out orderedContinuous	0
Number data bytes transferred orderedContinuous	0
Number data bytes transferred in orderedContinuous	0
Number data bytes transferred out orderedContinuous	0
Fragmented packets rate sortedContinuous	0
Bad fragment rate sortedContinuous	0
Max Src Window orderedContinuous	0
Max Dst Window orderedContinuous	0
Urgent rate sortedContinuous	0
Resend rate sortedContinuous	0
Wrong resend rate sortedContinuous	0
Duplicate ACK rate sortedContinuous	0
Wrong ACK sortedContinuous	0
Wrong data packet size rate sortedContinuous	0
Window exceeded rate sortedContinuous	0
Continued on next page	

Table L.2 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0
Number connection errors orderedContinuous	0
Number reset connection orderedContinuous	0
Number other errors orderedContinuous	0
Number disconnection errors orderedContinuous	0
Packet Destination IP nonKeyedSortedContinuous	0
Bytes Destination IP nonKeyedSortedContinuous	0
Connection Destination IP nonKeyedSortedContinuous	0
Packet Source IP nonKeyedSortedContinuous	0
Bytes Source IP nonKeyedSortedContinuous	0
Connection Source IP nonKeyedSortedContinuous	0

Table L.3: Similarity values for individual metrics of basecase 1,
train pair 3

Metric	Similarity
Packets in count	0.504219135401792
Connections in count	0.890322580645161
Bytes in count	0.212405591905261
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	0.11200771559948
Bytes Service discrete	0.136288061787165
Connection Service discrete	0.424300118318572
Packet Source port discrete	0.0683397315078339
Bytes Source port discrete	0.0364580839055527
Connection Source port discrete	0
Continued on next page	

Table L.3 – continued from previous page

Metric	Similarity
Connection Source port orderedContinuous	0
Packet TTL discrete	0.140439978948964
Packet TTL orderedContinuous	0.304262071543878
InterPacket delta sortedContinuous	0.3774081240654546
Packet sec orderedContinuous	0.968403099694479
Packet min orderedContinuous	0.777642587077063
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.4166666666666667
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.952071143556678
Bytes min orderedContinuous	0.525051961434484
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.4166666666666667
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.429779839549116
Packets in last w secs orderedContinuous	0.0990645552376704
Priv packets time rate sortedContinuous	0.9412394838294869
Unpriv packets time rate sortedContinuous	0.9933693593193471
Connections time rate sortedContinuous	0.5131978022821821
Priv connections connection time rate sortedContinuous	0.8311137769872555
Unpriv connections connection time rate sortedContinuous	0.9351162342224081
Priv packets priv connection time rate sortedContinuous	0.9444113298616622
Unpriv packets unpriv connection time rate sortedContinuous	0.4939724059500963
SYNs connection time rate sortedContinuous	0.9002917354558352
RSTs connection time rate sortedContinuous	0.9704378835635466
FINs connection time rate sortedContinuous	0.9627380457535541
PSH connection time rate sortedContinuous	0.565662017359871
Establishment errors connection time rate sortedContinuous	0.958337216284723
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.9531231222291798
Continued on next page	

Table L.3 – continued from previous page

Metric	Similarity
Priv packets packet rate sortedContinuous	0.9837832597623898
Unpriv packets packet rate sortedContinuous	0.9939739076794419
InterConnection delta sortedContinuous	0.5916537712663622
Connection sec orderedContinuous	0.600238927552802
Connection min orderedContinuous	0.201623562238328
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.4166666666666667
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.6738489079201373
Connection Priv connections rate sortedContinuous	0.7076352101581913
Connection Unpriv connections rate sortedContinuous	0.6355496707818224
Connection Priv packet rate sortedContinuous	0.4270540526588349
Connection Unpriv packet rate sortedContinuous	0.6909880922929329
Connection SYNs rate sortedContinuous	0.1086887437466695
Connection RSTs rate sortedContinuous	0.69758078630124
Connection FINs rate sortedContinuous	0.7422661877584741
Connection PSH rate sortedContinuous	0.7231603874622287
Connection Establishment errors rate sortedContinuous	0.4091491669131421
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.4069095864755017
Number of packets orderedContinuous	0.372316909312329
Number of packets in orderedContinuous	0.372316909312329
Number of packets out orderedContinuous	0.982676004423148
Duration sortedContinuous	0.8225806451612904
Number control packets rate sortedContinuous	0.7580645161290323
Number data packets rate sortedContinuous	0.7726426799007444
Number bytes transferred orderedContinuous	0.200399560372977
Number bytes transferred in orderedContinuous	0.200399560372977
Number bytes transferred out orderedContinuous	0.982676004423148
Number data bytes transferred orderedContinuous	0.140464979399647
Continued on next page	

Table L.3 – continued from previous page

Metric	Similarity
Number data bytes transferred in orderedContinuous	0.140464979399647
Number data bytes transferred out orderedContinuous	0.982676004423148
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0
Max Dst Window orderedContinuous	0
Urgent rate sortedContinuous	0
Resend rate sortedContinuous	0
Wrong resend rate sortedContinuous	0
Duplicate ACK rate sortedContinuous	0
Wrong ACK sortedContinuous	0
Wrong data packet size rate sortedContinuous	0
Window exceeded rate sortedContinuous	0
Hole rate sortedContinuous	0
Number connection errors orderedContinuous	0
Number reset connection orderedContinuous	0
Number other errors orderedContinuous	0
Number disconnection errors orderedContinuous	0
Packet Destination IP nonKeyedSortedContinuous	0.622823024971291
Bytes Destination IP nonKeyedSortedContinuous	0.484251608441661
Connection Destination IP nonKeyedSortedContinuous	0.714458560193587
Packet Source IP nonKeyedSortedContinuous	0.633620333762125
Bytes Source IP nonKeyedSortedContinuous	0.513457474799792
Connection Source IP nonKeyedSortedContinuous	0.824181348367395

Table L.4: Similarity values for individual metrics of basecase 1,
test pair 1

Metric	Similarity
Packets in count	0.623360780725831
Packets out count	0.639418710263397
Connections in count	0.988571428571429
Connections out count	0.895238095238095
Bytes in count	0.268525488951374
Bytes out count	0.586477051846563
SYN-ONLY rate ratio	0.896302975653742
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.87012404722762
Packet Service discrete	0.0709417134555924
Bytes Service discrete	0.0387320952090917
Connection Service discrete	0.196554185598704
Packet Source port discrete	0.0283365775366844
Bytes Source port discrete	0.016467746144173
Connection Source port discrete	0.032043013048081
Connection Source port orderedContinuous	0.0458208987385531
Packet TTL discrete	0.465200928340299
Packet TTL orderedContinuous	0.597513765069127
InterPacket delta sortedContinuous	0.5768541652975885
Packet sec orderedContinuous	0.59605503775442
Packet min orderedContinuous	0.601871268687709
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.4166666666666667
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.465710781049419
Bytes min orderedContinuous	0.285579239741244
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.4166666666666667
Continued on next page	

Table L.4 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.213027833655781
Packets in last w secs orderedContinuous	0.0930604859585746
Priv packets time rate sortedContinuous	0.9126617715485058
Unpriv packets time rate sortedContinuous	0.9464679876039514
Connections time rate sortedContinuous	0.5156586159444951
Priv connections connection time rate sortedContinuous	0.855935037174006
Unpriv connections connection time rate sortedContinuous	0.7132663694336415
Priv packets priv connection time rate sortedContinuous	0.5779200548109128
Unpriv packets unpriv connection time rate sortedContinuous	0.6131184299705099
SYNs connection time rate sortedContinuous	0.7812895855293393
RSTs connection time rate sortedContinuous	0.8723978580800944
FINs connection time rate sortedContinuous	0.4879446614856216
PSH connection time rate sortedContinuous	0.805995184404215
Establishment errors connection time rate sortedContinuous	0.8850714549638342
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.3786691273225609
Priv packets packet rate sortedContinuous	0.9794088510136283
Unpriv packets packet rate sortedContinuous	0.9739735008216136
InterConnection delta sortedContinuous	0.9201935771442604
Connection sec orderedContinuous	0.78303502597751
Connection min orderedContinuous	0.904828201395312
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.4166666666666667
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.8565218700450591
Connection Priv connections rate sortedContinuous	0.9533122933751091
Connection Unpriv connections rate sortedContinuous	0.9232073672951404
Connection Priv packet rate sortedContinuous	0.8571176219523562
Connection Unpriv packet rate sortedContinuous	0.9611192718233382
Continued on next page	

Table L.4 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.9299264762328011
Connection RSTs rate sortedContinuous	0.9383350151539672
Connection FINs rate sortedContinuous	0.8867557948051051
Connection PSH rate sortedContinuous	0.9091658845462029
Connection Establishment errors rate sortedContinuous	0.9407625647959534
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.8174413691121458
Number of packets orderedContinuous	0.240799019258682
Number of packets in orderedContinuous	0.251569209736419
Number of packets out orderedContinuous	0.278398195579913
Duration sortedContinuous	0.9098528213980355
Number control packets rate sortedContinuous	0.9695956483103863
Number data packets rate sortedContinuous	0.987478989797312
Number bytes transferred orderedContinuous	0.432066679734487
Number bytes transferred in orderedContinuous	0.406861848869731
Number bytes transferred out orderedContinuous	0.333979286579558
Number data bytes transferred orderedContinuous	0.211127607963238
Number data bytes transferred in orderedContinuous	0.142212942526781
Number data bytes transferred out orderedContinuous	0.212830221694501
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.589431623931624
Max Dst Window orderedContinuous	0.461185432924563
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9436619718309859
Wrong resend rate sortedContinuous	0.9577464788732394
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9859154929577465
Continued on next page	

Table L.4 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.886402188522528
Number reset connection orderedContinuous	0.931409208209698
Number other errors orderedContinuous	0.922009887109865
Number disconnection errors orderedContinuous	0.922009887109865
Packet Destination IP nonKeyedSortedContinuous	0.841594331812013
Bytes Destination IP nonKeyedSortedContinuous	0.827997351351257
Connection Destination IP nonKeyedSortedContinuous	0.873971975372866
Packet Source IP nonKeyedSortedContinuous	0.88202111260089
Bytes Source IP nonKeyedSortedContinuous	0.877305582388809
Connection Source IP nonKeyedSortedContinuous	0.952762690451909

Table L.5: Similarity values for individual metrics of basecase 1,
test pair 2

Metric	Similarity
Packets in count	0.888718481306154
Packets out count	0.842199200713925
Connections in count	0.673400009853673
Connections out count	0.632536815782162
Bytes in count	0.960632414757699
Bytes out count	0.530487715632451
SYN-ONLY rate ratio	0.490029774587402
SYN-ACK rate ratio	0.233381789348487
Idle connection rate ratio	0.465160571370682
Half-open connection rate ratio	0.967852238648114
Packet Service discrete	0.0210177664065308
Bytes Service discrete	0.0180480063954574
Connection Service discrete	0.021470932317585
Continued on next page	

Table L.5 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.0222271735023632
Bytes Source port discrete	0.0210315866303818
Connection Source port discrete	0.0180747374596897
Connection Source port orderedContinuous	0.0335036194646151
Packet TTL discrete	0.260483928944568
Packet TTL orderedContinuous	0.460938091437892
InterPacket delta sortedContinuous	0.92922071185118
Packet sec orderedContinuous	0.968882609340901
Packet min orderedContinuous	0.826617887993678
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.4166666666666667
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.951061351640612
Bytes min orderedContinuous	0.729076357517122
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.4166666666666667
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.58141359844495
Packets in last w secs orderedContinuous	0.525166415648127
Priv packets time rate sortedContinuous	0.5154104470661082
Unpriv packets time rate sortedContinuous	0.7537419799061425
Connections time rate sortedContinuous	0.5977081145755194
Priv connections connection time rate sortedContinuous	0.8102642887890632
Unpriv connections connection time rate sortedContinuous	0.699500132730458
Priv packets priv connection time rate sortedContinuous	0.6207693896769096
Unpriv packets unpriv connection time rate sortedContinuous	0.6116909358909754
SYNs connection time rate sortedContinuous	0.6458235907431483
RSTs connection time rate sortedContinuous	0.6250777033901452
FINs connection time rate sortedContinuous	0.7049746922039845
PSH connection time rate sortedContinuous	0.5338298646092606
Establishment errors connection time rate sortedContinuous	0.2691407181392873
Continued on next page	

Table L.5 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9682983031691531
Ave duration over last w secs sortedContinuous	0.2132005136277005
Priv packets packet rate sortedContinuous	0.4543892997456535
Unpriv packets packet rate sortedContinuous	0.7311779209815205
InterConnection delta sortedContinuous	0.859480549908986
Connection sec orderedContinuous	0.942054195285507
Connection min orderedContinuous	0.451138998350589
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.4166666666666667
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.3861267377469543
Connection Priv connections rate sortedContinuous	0.5805170395867459
Connection Unpriv connections rate sortedContinuous	0.4985937647168396
Connection Priv packet rate sortedContinuous	0.2076238018645213
Connection Unpriv packet rate sortedContinuous	0.6332395766469418
Connection SYNs rate sortedContinuous	0.6491827301210958
Connection RSTs rate sortedContinuous	0.9756901746531905
Connection FINs rate sortedContinuous	0.7897332748454488
Connection PSH rate sortedContinuous	0.4231030731563019
Connection Establishment errors rate sortedContinuous	0.7535614725397698
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9979995364779644
Ave duration over last m connections sortedContinuous	0.4978635188787819
Number of packets orderedContinuous	0.29245175209808
Number of packets in orderedContinuous	0.30883587766085
Number of packets out orderedContinuous	0.308325075340296
Duration sortedContinuous	0.981310265068085
Number control packets rate sortedContinuous	0.9894824667976781
Number data packets rate sortedContinuous	0.989450702943631
Number bytes transferred orderedContinuous	0.152168507411582
Continued on next page	

Table L.5 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.134813095596564
Number bytes transferred out orderedContinuous	0.245490983231689
Number data bytes transferred orderedContinuous	0.0963400662232439
Number data bytes transferred in orderedContinuous	0.0927369656803055
Number data bytes transferred out orderedContinuous	0.199176385779334
Fragmented packets rate sortedContinuous	0.9435688305962746
Bad fragment rate sortedContinuous	0.9435688305962746
Max Src Window orderedContinuous	0.101538923260219
Max Dst Window orderedContinuous	0.163074216121425
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.8733491625546387
Wrong resend rate sortedContinuous	0.9376154493151312
Duplicate ACK rate sortedContinuous	0.9977349943374858
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.942830997210196
Hole rate sortedContinuous	0.996823840505743
Number connection errors orderedContinuous	0.662907592463788
Number reset connection orderedContinuous	0.920293493252719
Number other errors orderedContinuous	0.900348953239168
Number disconnection errors orderedContinuous	0.670546433770186
Packet Destination IP nonKeyedSortedContinuous	0.833834992543772
Bytes Destination IP nonKeyedSortedContinuous	0.900180020122635
Connection Destination IP nonKeyedSortedContinuous	0.877253428679659
Packet Source IP nonKeyedSortedContinuous	0.767922678188645
Bytes Source IP nonKeyedSortedContinuous	0.748186952944122
Connection Source IP nonKeyedSortedContinuous	0.736309179088121

Table L.6: Similarity values for individual metrics of basecase 1,
test pair 3

Metric	Similarity
Packets in count	0.924834749763928
Connections in count	0.571428571428571
Bytes in count	0.958228404754579
SYN-ONLY rate ratio	1
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	1
Packet Service discrete	0.0488961420246148
Bytes Service discrete	0.0320822862474621
Connection Service discrete	0.0960046897546898
Packet Source port discrete	0.0154296714752376
Bytes Source port discrete	0.0148745157126445
Connection Source port discrete	0.00420168067226891
Connection Source port orderedContinuous	0.0144521456631196
Packet TTL discrete	0.189416831939512
Packet TTL orderedContinuous	0.397030736657087
InterPacket delta sortedContinuous	0.7576545085710267
Packet sec orderedContinuous	0.786462461399049
Packet min orderedContinuous	0.504589008141233
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.4166666666666667
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.770047795173831
Bytes min orderedContinuous	0.460673798539426
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.4166666666666667
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.643770325880944
Packets in last w secs orderedContinuous	0.230237262786133
Continued on next page	

Table L.6 – continued from previous page

Metric	Similarity
Priv packets time rate sortedContinuous	0.9997957933428629
Unpriv packets time rate sortedContinuous	0.9991857003043262
Connections time rate sortedContinuous	0.8161515821421305
Priv connections connection time rate sortedContinuous	0.8798603195222537
Unpriv connections connection time rate sortedContinuous	0.4416792392770382
Priv packets priv connection time rate sortedContinuous	0.9997957933428629
Unpriv packets unpriv connection time rate sortedContinuous	0.4107176764438019
SYNs connection time rate sortedContinuous	0.9426921051213387
RSTs connection time rate sortedContinuous	0.9180926434394274
FINs connection time rate sortedContinuous	0.8729958789783064
PSH connection time rate sortedContinuous	0.8671063388054073
Establishment errors connection time rate sortedContinuous	0.6845064192862692
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.6434985768512852
Priv packets packet rate sortedContinuous	0.997004969028657
Unpriv packets packet rate sortedContinuous	0.999529271581245
InterConnection delta sortedContinuous	0.6160161707096008
Connection sec orderedContinuous	0.728036873252392
Connection min orderedContinuous	0.63124705099219
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.4166666666666667
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.7656529294502449
Connection Priv connections rate sortedContinuous	0.8334446421716334
Connection Unpriv connections rate sortedContinuous	0.5184368179996057
Connection Priv packet rate sortedContinuous	0.7682651397196475
Connection Unpriv packet rate sortedContinuous	0.5390952481367238
Connection SYNs rate sortedContinuous	0.974595114522526
Connection RSTs rate sortedContinuous	0.8558062620538334
Connection FINs rate sortedContinuous	0.9719985870523628
Continued on next page	

Table L.6 – continued from previous page

Metric	Similarity
Connection PSH rate sortedContinuous	0.8406866816639747
Connection Establishment errors rate sortedContinuous	0.5209802649570654
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.4320674219692712
Number of packets orderedContinuous	0.399108783374721
Number of packets in orderedContinuous	0.399108783374721
Number of packets out orderedContinuous	0.983957219251337
Duration sortedContinuous	0.604799135112256
Number control packets rate sortedContinuous	0.9256207752940556
Number data packets rate sortedContinuous	0.7228070175438596
Number bytes transferred orderedContinuous	0.0845690392849704
Number bytes transferred in orderedContinuous	0.0845690392849704
Number bytes transferred out orderedContinuous	0.983957219251337
Number data bytes transferred orderedContinuous	0.0360229001869869
Number data bytes transferred in orderedContinuous	0.0360229001869869
Number data bytes transferred out orderedContinuous	0.983957219251337
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.449044674688057
Max Dst Window orderedContinuous	0.557377049180328
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	1
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.5294117647058824
Window exceeded rate sortedContinuous	1
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.359488633503396
Number reset connection orderedContinuous	0.982658959537572
Continued on next page	

Table L.6 – continued from previous page

Metric	Similarity
Number other errors orderedContinuous	0.982658959537572
Number disconnection errors orderedContinuous	0.982658959537572
Packet Destination IP nonKeyedSortedContinuous	0.716451231897129
Bytes Destination IP nonKeyedSortedContinuous	0.601815823836498
Connection Destination IP nonKeyedSortedContinuous	0.84036683904331
Packet Source IP nonKeyedSortedContinuous	0.585759047787207
Bytes Source IP nonKeyedSortedContinuous	0.523308787240186
Connection Source IP nonKeyedSortedContinuous	0.645610469047067

Table L.7: Similarity values for individual metrics of basecase 2,
train pair 1

Metric	Similarity
Packets in count	0.459016393442623
Packets out count	0.389121338912134
Connections in count	0.558704453441295
Connections out count	0.459459459459459
Bytes in count	0.336624136048631
Bytes out count	0.304028790428117
SYN-ONLY rate ratio	0.602739726027395
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.602739726027395
Packet Service discrete	0.0253313368112803
Bytes Service discrete	0.0225393991472132
Connection Service discrete	0.0292044836386599
Packet Source port discrete	0.0208590719103628
Bytes Source port discrete	0.0183208134491122
Connection Source port discrete	0.0156904836364318
Continued on next page	

Table L.7 – continued from previous page

Metric	Similarity
Connection Source port orderedContinuous	0.0436806584407237
Packet TTL discrete	0.231765702738033
Packet TTL orderedContinuous	0.45437944869474
InterPacket delta sortedContinuous	0.6029075962416924
Packet sec orderedContinuous	0.532140829896962
Packet min orderedContinuous	0.121208686446773
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.4166666666666667
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.443495596383846
Bytes min orderedContinuous	0.100063128120056
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.4166666666666667
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.303796050458534
Packets in last w secs orderedContinuous	0.299111792108158
Priv packets time rate sortedContinuous	0.9065462394101329
Unpriv packets time rate sortedContinuous	0.8920763745112638
Connections time rate sortedContinuous	0.7026627176677567
Priv connections connection time rate sortedContinuous	0.9297560975609756
Unpriv connections connection time rate sortedContinuous	0.9739837398373983
Priv packets priv connection time rate sortedContinuous	0.7648525436760168
Unpriv packets unpriv connection time rate sortedContinuous	0.9739837398373983
SYNs connection time rate sortedContinuous	0.895085904377981
RSTs connection time rate sortedContinuous	0.9488464958054624
FINs connection time rate sortedContinuous	0.7104611989505622
PSH connection time rate sortedContinuous	0.66492996865027
Establishment errors connection time rate sortedContinuous	0.9853658536585366
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.7554535619185778
Continued on next page	

Table L.7 – continued from previous page

Metric	Similarity
Priv packets packet rate sortedContinuous	0.9620227543728355
Unpriv packets packet rate sortedContinuous	0.9716222575716968
InterConnection delta sortedContinuous	0.5999927279376213
Connection sec orderedContinuous	0.467041581908104
Connection min orderedContinuous	0.108972932969077
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.4166666666666667
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.9074438723881759
Connection Priv connections rate sortedContinuous	0.9304768370342303
Connection Unpriv connections rate sortedContinuous	0.6115940754741379
Connection Priv packet rate sortedContinuous	0.9264954809554058
Connection Unpriv packet rate sortedContinuous	0.7919534632034631
Connection SYNs rate sortedContinuous	0.6420914941775731
Connection RSTs rate sortedContinuous	0.5537156814570612
Connection FINs rate sortedContinuous	0.6934392872865832
Connection PSH rate sortedContinuous	0.706920098523017
Connection Establishment errors rate sortedContinuous	0.4666666666666667
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.7447751404782913
Number of packets orderedContinuous	0.305776745778835
Number of packets in orderedContinuous	0.462073243066722
Number of packets out orderedContinuous	0.447673780462245
Duration sortedContinuous	0.9417635691346366
Number control packets rate sortedContinuous	0.9428487449490981
Number data packets rate sortedContinuous	0.9334804758533572
Number bytes transferred orderedContinuous	0.258446839258944
Number bytes transferred in orderedContinuous	0.20796828578627
Number bytes transferred out orderedContinuous	0.213710056878294
Number data bytes transferred orderedContinuous	0.0458724422462858
Continued on next page	

Table L.7 – continued from previous page

Metric	Similarity
Number data bytes transferred in orderedContinuous	0.0512064471586446
Number data bytes transferred out orderedContinuous	0.262384589530796
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.225319396051103
Max Dst Window orderedContinuous	0.145098039215686
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.8181818181818182
Wrong resend rate sortedContinuous	0.9498432601880877
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.8450074515648286
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.762267724075325
Number reset connection orderedContinuous	0.669556472727292
Number other errors orderedContinuous	0.693995577479163
Number disconnection errors orderedContinuous	0.693995577479163
Packet Destination IP nonKeyedSortedContinuous	0.780577499081161
Bytes Destination IP nonKeyedSortedContinuous	0.723509154956813
Connection Destination IP nonKeyedSortedContinuous	0.298582995951417
Packet Source IP nonKeyedSortedContinuous	0.645957321799819
Bytes Source IP nonKeyedSortedContinuous	0.460739739864239
Connection Source IP nonKeyedSortedContinuous	0.849457832216453

Table L.8: Similarity values for individual metrics of basemode 2,
train pair 2

Metric	Similarity
Packets in count	0
Packets out count	0
Connections in count	0
Connections out count	0
Bytes in count	0
Bytes out count	0
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	0
Bytes Service discrete	0
Connection Service discrete	0
Packet Source port discrete	0
Bytes Source port discrete	0
Connection Source port discrete	0
Connection Source port orderedContinuous	0
Packet TTL discrete	0
Packet TTL orderedContinuous	0
InterPacket delta sortedContinuous	0
Packet sec orderedContinuous	0
Packet min orderedContinuous	0
Packet GmHour orderedContinuous	0
Packet LocHour orderedContinuous	0
Packet weekday orderedContinuous	0
Bytes sec orderedContinuous	0
Bytes min orderedContinuous	0
Bytes GmHour orderedContinuous	0
Bytes LocHour orderedContinuous	0
Continued on next page	

Table L.8 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	0
Packet size orderedContinuous	0
Packets in last w secs orderedContinuous	0
Priv packets time rate sortedContinuous	0
Unpriv packets time rate sortedContinuous	0
Connections time rate sortedContinuous	0
Priv connections connection time rate sortedContinuous	0
Unpriv connections connection time rate sortedContinuous	0
Priv packets priv connection time rate sortedContinuous	0
Unpriv packets unpriv connection time rate sortedContinuous	0
SYNs connection time rate sortedContinuous	0
RSTs connection time rate sortedContinuous	0
FINs connection time rate sortedContinuous	0
PSH connection time rate sortedContinuous	0
Establishment errors connection time rate sortedContinuous	0
Other errors connection time rate sortedContinuous	0
Disconnection errors connection time rate sortedContinuous	0
Ave duration over last w secs sortedContinuous	0
Priv packets packet rate sortedContinuous	0
Unpriv packets packet rate sortedContinuous	0
InterConnection delta sortedContinuous	0
Connection sec orderedContinuous	0
Connection min orderedContinuous	0
Connection GmHour orderedContinuous	0
Connection LocHour orderedContinuous	0
Connection weekday orderedContinuous	0
Connection packet rate sortedContinuous	0
Connection Priv connections rate sortedContinuous	0
Connection Unpriv connections rate sortedContinuous	0
Connection Priv packet rate sortedContinuous	0
Connection Unpriv packet rate sortedContinuous	0
Continued on next page	

Table L.8 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0
Connection RSTs rate sortedContinuous	0
Connection FINs rate sortedContinuous	0
Connection PSH rate sortedContinuous	0
Connection Establishment errors rate sortedContinuous	0
Connection Other errors rate sortedContinuous	0
Connection Disconnection errors rate sortedContinuous	0
Ave duration over last m connections sortedContinuous	0
Number of packets orderedContinuous	0
Number of packets in orderedContinuous	0
Number of packets out orderedContinuous	0
Duration sortedContinuous	0
Number control packets rate sortedContinuous	0
Number data packets rate sortedContinuous	0
Number bytes transferred orderedContinuous	0
Number bytes transferred in orderedContinuous	0
Number bytes transferred out orderedContinuous	0
Number data bytes transferred orderedContinuous	0
Number data bytes transferred in orderedContinuous	0
Number data bytes transferred out orderedContinuous	0
Fragmented packets rate sortedContinuous	0
Bad fragment rate sortedContinuous	0
Max Src Window orderedContinuous	0
Max Dst Window orderedContinuous	0
Urgent rate sortedContinuous	0
Resend rate sortedContinuous	0
Wrong resend rate sortedContinuous	0
Duplicate ACK rate sortedContinuous	0
Wrong ACK sortedContinuous	0
Wrong data packet size rate sortedContinuous	0
Window exceeded rate sortedContinuous	0
Continued on next page	

Table L.8 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0
Number connection errors orderedContinuous	0
Number reset connection orderedContinuous	0
Number other errors orderedContinuous	0
Number disconnection errors orderedContinuous	0
Packet Destination IP nonKeyedSortedContinuous	0
Bytes Destination IP nonKeyedSortedContinuous	0
Connection Destination IP nonKeyedSortedContinuous	0
Packet Source IP nonKeyedSortedContinuous	0
Bytes Source IP nonKeyedSortedContinuous	0
Connection Source IP nonKeyedSortedContinuous	0

Table L.9: Similarity values for individual metrics of basecase 2,
train pair 3

Metric	Similarity
Packets in count	0.966150616636054
Packets out count	0.974617159639186
Connections in count	0.952775476387738
Connections out count	0.933135215453195
Bytes in count	0.931182290320306
Bytes out count	0.928956818757064
SYN-ONLY rate ratio	0.540644805731607
SYN-ACK rate ratio	1
Idle connection rate ratio	0.993920972644377
Half-open connection rate ratio	0.700339104051403
Packet Service discrete	0.0473882418841146
Bytes Service discrete	0.0462576587917518
Connection Service discrete	0.0823271036349928
Continued on next page	

Table L.9 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.0399985115018171
Bytes Source port discrete	0.0390099628283373
Connection Source port discrete	0.0171788543329378
Connection Source port orderedContinuous	0.0432396859773346
Packet TTL discrete	0.732648382458296
Packet TTL orderedContinuous	0.720585999257884
InterPacket delta sortedContinuous	0.9348414827448465
Packet sec orderedContinuous	0.730447799129595
Packet min orderedContinuous	0.912638816825567
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.5
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.76255212051161
Bytes min orderedContinuous	0.880341821462971
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.5
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.591514825239273
Packets in last w secs orderedContinuous	0.708390696452823
Priv packets time rate sortedContinuous	0.9777025477443704
Unpriv packets time rate sortedContinuous	0.9899545316497203
Connections time rate sortedContinuous	0.9792456481216251
Priv connections connection time rate sortedContinuous	0.9827908662214419
Unpriv connections connection time rate sortedContinuous	0.9772297111858882
Priv packets priv connection time rate sortedContinuous	0.9790363360200305
Unpriv packets unpriv connection time rate sortedContinuous	0.978788591238055
SYNs connection time rate sortedContinuous	0.9473604593462875
RSTs connection time rate sortedContinuous	0.9588190318717774
FINs connection time rate sortedContinuous	0.9199120771778432
PSH connection time rate sortedContinuous	0.9744725199064435
Establishment errors connection time rate sortedContinuous	0.9525468965113986
Continued on next page	

Table L.9 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.900082236851798
Priv packets packet rate sortedContinuous	0.9820121543761438
Unpriv packets packet rate sortedContinuous	0.9839381560062615
InterConnection delta sortedContinuous	0.9223028069785719
Connection sec orderedContinuous	0.825178975291632
Connection min orderedContinuous	0.829228015744727
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.5
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.9774923025438553
Connection Priv connections rate sortedContinuous	0.9614456912038386
Connection Unpriv connections rate sortedContinuous	0.9333527007882912
Connection Priv packet rate sortedContinuous	0.9804175818652203
Connection Unpriv packet rate sortedContinuous	0.9153679648479271
Connection SYNs rate sortedContinuous	0.9703866641890447
Connection RSTs rate sortedContinuous	0.916053408352031
Connection FINs rate sortedContinuous	0.9681049879513663
Connection PSH rate sortedContinuous	0.9647609294742062
Connection Establishment errors rate sortedContinuous	0.8932871539149242
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.817026365678302
Number of packets orderedContinuous	0.646220969407005
Number of packets in orderedContinuous	0.906204336633778
Number of packets out orderedContinuous	0.729897442886173
Duration sortedContinuous	0.9443071327510204
Number control packets rate sortedContinuous	0.9456016593109618
Number data packets rate sortedContinuous	0.9343125061287311
Number bytes transferred orderedContinuous	0.528475878542352
Continued on next page	

Table L.9 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.479154047078997
Number bytes transferred out orderedContinuous	0.561488235582825
Number data bytes transferred orderedContinuous	0.360302098448058
Number data bytes transferred in orderedContinuous	0.384890655896899
Number data bytes transferred out orderedContinuous	0.419851554065178
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.618996491723764
Max Dst Window orderedContinuous	0.805281400357801
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9969418960244648
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	1
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.868622233459544
Number reset connection orderedContinuous	0.951221144080537
Number other errors orderedContinuous	0.951807813480486
Number disconnection errors orderedContinuous	0.951807813480486
Packet Destination IP nonKeyedSortedContinuous	0.953984955208972
Bytes Destination IP nonKeyedSortedContinuous	0.945055648435653
Connection Destination IP nonKeyedSortedContinuous	0.938745239207946
Packet Source IP nonKeyedSortedContinuous	0.948105109197426
Bytes Source IP nonKeyedSortedContinuous	0.934980329704157
Connection Source IP nonKeyedSortedContinuous	0.953352613539774

Table L.10: Similarity values for individual metrics of basecase 2,
test pair 1

Metric	Similarity
Packets in count	0.520334583213153
Packets out count	0.475601821730644
Connections in count	0.878048780487805
Connections out count	0.567811934900542
Bytes in count	0.287290295499875
Bytes out count	0.308037971688126
SYN-ONLY rate ratio	0.7527352297593
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.883994126284875
Packet Service discrete	0.0219343651486217
Bytes Service discrete	0.012678523649078
Connection Service discrete	0.0736726142490965
Packet Source port discrete	0.0138069967839507
Bytes Source port discrete	0.00924110247434488
Connection Source port discrete	0.0161615068062736
Connection Source port orderedContinuous	0.0376105636723555
Packet TTL discrete	0.253949892379562
Packet TTL orderedContinuous	0.355539195446117
InterPacket delta sortedContinuous	0.6082499618974513
Packet sec orderedContinuous	0.529676261435055
Packet min orderedContinuous	0.501489830414914
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.4166666666666667
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.467156358183783
Bytes min orderedContinuous	0.302953330488676
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.4166666666666667
Continued on next page	

Table L.10 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.127818640686501
Packets in last w secs orderedContinuous	0.106033304134028
Priv packets time rate sortedContinuous	0.9092523707761761
Unpriv packets time rate sortedContinuous	0.9491361284647234
Connections time rate sortedContinuous	0.6166643611635944
Priv connections connection time rate sortedContinuous	0.8391596009469296
Unpriv connections connection time rate sortedContinuous	0.7351191667636484
Priv packets priv connection time rate sortedContinuous	0.6920694539628171
Unpriv packets unpriv connection time rate sortedContinuous	0.691110311350501
SYNs connection time rate sortedContinuous	0.5177974240997133
RSTs connection time rate sortedContinuous	0.8648512801399605
FINs connection time rate sortedContinuous	0.4310129250640362
PSH connection time rate sortedContinuous	0.7300369240087585
Establishment errors connection time rate sortedContinuous	0.9823673645884048
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.3217664169592307
Priv packets packet rate sortedContinuous	0.9681764446343309
Unpriv packets packet rate sortedContinuous	0.9797851299199163
InterConnection delta sortedContinuous	0.6902438067946578
Connection sec orderedContinuous	0.728785230012793
Connection min orderedContinuous	0.696630067328683
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.4166666666666667
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.7507192960647773
Connection Priv connections rate sortedContinuous	0.9161080823414112
Connection Unpriv connections rate sortedContinuous	0.4874091541494399
Connection Priv packet rate sortedContinuous	0.7983038057510277
Connection Unpriv packet rate sortedContinuous	0.7720285721103428
Continued on next page	

Table L.10 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.5977945286413662
Connection RSTs rate sortedContinuous	0.6313404193259297
Connection FINs rate sortedContinuous	0.688504815568355
Connection PSH rate sortedContinuous	0.8007689722344749
Connection Establishment errors rate sortedContinuous	0.9431245500359971
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.4106485361697882
Number of packets orderedContinuous	0.21250042292504
Number of packets in orderedContinuous	0.256742461347006
Number of packets out orderedContinuous	0.343813351852284
Duration sortedContinuous	0.7718814207703705
Number control packets rate sortedContinuous	0.8394348419657443
Number data packets rate sortedContinuous	0.9163516657625706
Number bytes transferred orderedContinuous	0.125868897620604
Number bytes transferred in orderedContinuous	0.12721509138952
Number bytes transferred out orderedContinuous	0.0969652941568114
Number data bytes transferred orderedContinuous	0.0503160765164305
Number data bytes transferred in orderedContinuous	0.0317646825665711
Number data bytes transferred out orderedContinuous	0.0536306597405765
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.64577807250221
Max Dst Window orderedContinuous	0.0711787672084166
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.6016058439525084
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9892080863353094
Continued on next page	

Table L.10 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9767441860465116
Number connection errors orderedContinuous	0.401159461111988
Number reset connection orderedContinuous	0.613241375482375
Number other errors orderedContinuous	0.555631264828299
Number disconnection errors orderedContinuous	0.555631264828299
Packet Destination IP nonKeyedSortedContinuous	0.773110766964389
Bytes Destination IP nonKeyedSortedContinuous	0.779945877778592
Connection Destination IP nonKeyedSortedContinuous	0.744799445805929
Packet Source IP nonKeyedSortedContinuous	0.691873967462269
Bytes Source IP nonKeyedSortedContinuous	0.671842032544211
Connection Source IP nonKeyedSortedContinuous	0.913378796413441

Table L.11: Similarity values for individual metrics of basecase 2,
test pair 3

Metric	Similarity
Packets in count	0.933932007697242
Packets out count	0.876095118898623
Connections in count	0.892173913043478
Connections out count	0.821529745042493
Bytes in count	0.96509782527835
Bytes out count	0.918495378737027
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	0.38005289444008
Bytes Service discrete	0.392077119905808
Connection Service discrete	0.791413171557883
Continued on next page	

Table L.11 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.752449097616427
Bytes Source port discrete	0.775070332466676
Connection Source port discrete	0.755037054360465
Connection Source port orderedContinuous	0.713212524444874
Packet TTL discrete	0.713654684005721
Packet TTL orderedContinuous	0.670657952734159
InterPacket delta sortedContinuous	0.8927775349246756
Packet sec orderedContinuous	0.886788597957976
Packet min orderedContinuous	0.894436078604771
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.5
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.885687693602752
Bytes min orderedContinuous	0.893181850863739
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.5
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.775664626184659
Packets in last w secs orderedContinuous	0.599524159467745
Priv packets time rate sortedContinuous	0.9044673255381801
Unpriv packets time rate sortedContinuous	0.8878435455724562
Connections time rate sortedContinuous	0.9454900505565142
Priv connections connection time rate sortedContinuous	0.8583198130630561
Unpriv connections connection time rate sortedContinuous	0.910749613617048
Priv packets priv connection time rate sortedContinuous	0.9417464487075931
Unpriv packets unpriv connection time rate sortedContinuous	0.9350273548571999
SYNs connection time rate sortedContinuous	0.9311531192279654
RSTs connection time rate sortedContinuous	0.9935064935064936
FINs connection time rate sortedContinuous	0.9488881940077845
PSH connection time rate sortedContinuous	0.9000840622864906
Establishment errors connection time rate sortedContinuous	1
Continued on next page	

Table L.11 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.974025974025974
Priv packets packet rate sortedContinuous	0.7398070913412504
Unpriv packets packet rate sortedContinuous	0.8932230860181089
InterConnection delta sortedContinuous	0.8578816405460318
Connection sec orderedContinuous	0.880783913080375
Connection min orderedContinuous	0.880020208184971
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.5
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.9728730109093979
Connection Priv connections rate sortedContinuous	0.7655884416212395
Connection Unpriv connections rate sortedContinuous	0.8688253856776001
Connection Priv packet rate sortedContinuous	0.7706766345017969
Connection Unpriv packet rate sortedContinuous	0.9950186799501868
Connection SYNs rate sortedContinuous	0.9662812073770978
Connection RSTs rate sortedContinuous	0.9701120797011208
Connection FINs rate sortedContinuous	0.9385637193856371
Connection PSH rate sortedContinuous	0.9537573107626418
Connection Establishment errors rate sortedContinuous	1
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.9638854296388543
Number of packets orderedContinuous	0.499688473520249
Number of packets in orderedContinuous	0.756465256937679
Number of packets out orderedContinuous	0.559558407641097
Duration sortedContinuous	0.9987546699875467
Number control packets rate sortedContinuous	0.9987546699875467
Number data packets rate sortedContinuous	0.9994465199944652
Number bytes transferred orderedContinuous	0.621835293142756
Continued on next page	

Table L.11 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.678911668293287
Number bytes transferred out orderedContinuous	0.519328411108666
Number data bytes transferred orderedContinuous	0.499688473520249
Number data bytes transferred in orderedContinuous	1
Number data bytes transferred out orderedContinuous	0.499688473520249
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0
Max Dst Window orderedContinuous	0
Urgent rate sortedContinuous	0
Resend rate sortedContinuous	0
Wrong resend rate sortedContinuous	0
Duplicate ACK rate sortedContinuous	0
Wrong ACK sortedContinuous	0
Wrong data packet size rate sortedContinuous	0
Window exceeded rate sortedContinuous	0
Hole rate sortedContinuous	0
Number connection errors orderedContinuous	0
Number reset connection orderedContinuous	0
Number other errors orderedContinuous	0
Number disconnection errors orderedContinuous	0
Packet Destination IP nonKeyedSortedContinuous	0.920868791722713
Bytes Destination IP nonKeyedSortedContinuous	0.938358983421838
Connection Destination IP nonKeyedSortedContinuous	0.917702140535552
Packet Source IP nonKeyedSortedContinuous	0.925021960243768
Bytes Source IP nonKeyedSortedContinuous	0.924483348955672
Connection Source IP nonKeyedSortedContinuous	0.92032324131806

Table L.12: Similarity values for individual metrics of basecase 3,
train pair 1

Metric	Similarity
Packets in count	0.146128468853095
Packets out count	0.119143349470748
Connections in count	0.482142857142857
Connections out count	0.456273764258555
Bytes in count	0.294687281483518
Bytes out count	0.0111565995004193
SYN-ONLY rate ratio	0.24
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.428571428571429
Packet Service discrete	0.00163698382679438
Bytes Service discrete	0.000647724205547792
Connection Service discrete	0.00976271560904016
Packet Source port discrete	0.000968038153559763
Bytes Source port discrete	0.00039031005716315
Connection Source port discrete	0.00261237782968303
Connection Source port orderedContinuous	0.0087621569554315
Packet TTL discrete	0.0724482079863377
Packet TTL orderedContinuous	0.133387193675124
InterPacket delta sortedContinuous	0.5269465315387591
Packet sec orderedContinuous	0.738654431532183
Packet min orderedContinuous	0.185208649503984
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.655466183255257
Bytes min orderedContinuous	0.149186884260556
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Continued on next page	

Table L.12 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.108639328273
Packets in last w secs orderedContinuous	0.0693521838707461
Priv packets time rate sortedContinuous	0.9149536309069954
Unpriv packets time rate sortedContinuous	0.9100183196921223
Connections time rate sortedContinuous	0.05351234507888481
Priv connections connection time rate sortedContinuous	0.9912003254802443
Unpriv connections connection time rate sortedContinuous	0.9993238674780257
Priv packets priv connection time rate sortedContinuous	0.06583780671888467
Unpriv packets unpriv connection time rate sortedContinuous	0.9993238674780257
SYNs connection time rate sortedContinuous	0.07219134992281775
RSTs connection time rate sortedContinuous	0.7932945698398885
FINs connection time rate sortedContinuous	0.4341201732682463
PSH connection time rate sortedContinuous	0.03210420820055335
Establishment errors connection time rate sortedContinuous	0.9993238674780257
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.07947629468275516
Priv packets packet rate sortedContinuous	0.9436718471527309
Unpriv packets packet rate sortedContinuous	0.9391609003908199
InterConnection delta sortedContinuous	0.6995088260355733
Connection sec orderedContinuous	0.381681309200672
Connection min orderedContinuous	0.238853330340308
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.3419193425109644
Connection Priv connections rate sortedContinuous	0.9928294612361973
Connection Unpriv connections rate sortedContinuous	0.9372549019607843
Connection Priv packet rate sortedContinuous	0.3418811300449327
Connection Unpriv packet rate sortedContinuous	0.9372549019607843

Continued on next page

Table L.12 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.6702203698778888
Connection RSTs rate sortedContinuous	0.851569954461343
Connection FINs rate sortedContinuous	0.7668654379621871
Connection PSH rate sortedContinuous	0.30342143654388
Connection Establishment errors rate sortedContinuous	0.8085561497326203
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.688759906843497
Number of packets orderedContinuous	0.26397748635623
Number of packets in orderedContinuous	0.393384882890583
Number of packets out orderedContinuous	0.487725416618575
Duration sortedContinuous	0.9527566841018854
Number control packets rate sortedContinuous	0.959079513049362
Number data packets rate sortedContinuous	0.9731398047852885
Number bytes transferred orderedContinuous	0.245671840673684
Number bytes transferred in orderedContinuous	0.220140632059115
Number bytes transferred out orderedContinuous	0.161038523180906
Number data bytes transferred orderedContinuous	0.0319830580490429
Number data bytes transferred in orderedContinuous	0.0501784097442987
Number data bytes transferred out orderedContinuous	0.137397765197826
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.362091503267974
Max Dst Window orderedContinuous	0.125248015873016
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.7391387927720108
Wrong resend rate sortedContinuous	0.9444444444444444
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9265700483091788
Continued on next page	

Table L.12 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.608875313345266
Number reset connection orderedContinuous	0.579376095753356
Number other errors orderedContinuous	0.545507584597433
Number disconnection errors orderedContinuous	0.545507584597433
Packet Destination IP nonKeyedSortedContinuous	0.657000823608826
Bytes Destination IP nonKeyedSortedContinuous	0.633872699586934
Connection Destination IP nonKeyedSortedContinuous	0.565407509157509
Packet Source IP nonKeyedSortedContinuous	0.659530577833174
Bytes Source IP nonKeyedSortedContinuous	0.597940315950173
Connection Source IP nonKeyedSortedContinuous	0.925301361519908

Table L.13: Similarity values for individual metrics of basecase 3,
train pair 2

Metric	Similarity
Packets in count	0.195338585952351
Packets out count	0.305216519806303
Connections in count	0.633116215289681
Connections out count	0.542128069330766
Bytes in count	0.14662792183027
Bytes out count	0.0435434761366794
SYN-ONLY rate ratio	0.877266418018012
SYN-ACK rate ratio	0
Idle connection rate ratio	0.249713226859614
Half-open connection rate ratio	0.900374553500413
Packet Service discrete	0.0106697673164339
Bytes Service discrete	0.00712446539657812
Connection Service discrete	0.00930655616474892
Continued on next page	

Table L.13 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.015520658565767
Bytes Source port discrete	0.0125431374734062
Connection Source port discrete	0.0279220372326765
Connection Source port orderedContinuous	0.0456077168413771
Packet TTL discrete	0.134714754619112
Packet TTL orderedContinuous	0.20752774654715
InterPacket delta sortedContinuous	0.279050400652486
Packet sec orderedContinuous	0.951750147820615
Packet min orderedContinuous	0.660637416387861
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.940745332442758
Bytes min orderedContinuous	0.60807970096962
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.458606434672808
Packets in last w secs orderedContinuous	0.0452898014612499
Priv packets time rate sortedContinuous	0.07403157420275869
Unpriv packets time rate sortedContinuous	0.8490875779573812
Connections time rate sortedContinuous	0.3519171783114016
Priv connections connection time rate sortedContinuous	0.5731480507198193
Unpriv connections connection time rate sortedContinuous	0.4998566045956648
Priv packets priv connection time rate sortedContinuous	0.5167871260693338
Unpriv packets unpriv connection time rate sortedContinuous	0.6555883863082458
SYNs connection time rate sortedContinuous	0.6861598439916532
RSTs connection time rate sortedContinuous	0.6407719104948917
FINs connection time rate sortedContinuous	0.7337966893594424
PSH connection time rate sortedContinuous	0.5511830973948759
Establishment errors connection time rate sortedContinuous	0.6295292467263538
Continued on next page	

Table L.13 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9674811419496878
Ave duration over last w secs sortedContinuous	0.5783642726645121
Priv packets packet rate sortedContinuous	0.1425885612839585
Unpriv packets packet rate sortedContinuous	0.8352852438346972
InterConnection delta sortedContinuous	0.4450883792593257
Connection sec orderedContinuous	0.890301169287926
Connection min orderedContinuous	0.603407243518861
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.1960832730885282
Connection Priv connections rate sortedContinuous	0.4599305610847998
Connection Unpriv connections rate sortedContinuous	0.2275486718294338
Connection Priv packet rate sortedContinuous	0.7041151035255404
Connection Unpriv packet rate sortedContinuous	0.7970684338628934
Connection SYNs rate sortedContinuous	0.5264242038459971
Connection RSTs rate sortedContinuous	0.6297801279281129
Connection FINs rate sortedContinuous	0.7132913421967848
Connection PSH rate sortedContinuous	0.3881073385476067
Connection Establishment errors rate sortedContinuous	0.801964025819004
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9951560818083961
Ave duration over last m connections sortedContinuous	0.1261105941641809
Number of packets orderedContinuous	0.238225179890859
Number of packets in orderedContinuous	0.278422614915211
Number of packets out orderedContinuous	0.232092068300009
Duration sortedContinuous	0.6323459610256431
Number control packets rate sortedContinuous	0.793349619399785
Number data packets rate sortedContinuous	0.8335705532024947
Number bytes transferred orderedContinuous	0.086100373962204
Continued on next page	

Table L.13 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.0868695295094778
Number bytes transferred out orderedContinuous	0.174544300112142
Number data bytes transferred orderedContinuous	0.0545535469293516
Number data bytes transferred in orderedContinuous	0.0576647559602387
Number data bytes transferred out orderedContinuous	0.168931518532704
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.225920031184259
Max Dst Window orderedContinuous	0.0846525981342349
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9553682747174872
Wrong resend rate sortedContinuous	0.9693007551723984
Duplicate ACK rate sortedContinuous	0.985940437623629
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9848056537102473
Window exceeded rate sortedContinuous	0.9943049901639988
Hole rate sortedContinuous	0.97615945389319
Number connection errors orderedContinuous	0.392658141342142
Number reset connection orderedContinuous	0.669476937505875
Number other errors orderedContinuous	0.662244942679469
Number disconnection errors orderedContinuous	0.777257802167677
Packet Destination IP nonKeyedSortedContinuous	0.741794331917233
Bytes Destination IP nonKeyedSortedContinuous	0.617430342109565
Connection Destination IP nonKeyedSortedContinuous	0.873339260121403
Packet Source IP nonKeyedSortedContinuous	0.757735872416305
Bytes Source IP nonKeyedSortedContinuous	0.6462773821466
Connection Source IP nonKeyedSortedContinuous	0.548786831277194

Table L.14: Similarity values for individual metrics of basecase 3,
train pair 3

Metric	Similarity
Packets in count	0.839974118377155
Packets out count	0.817847184180343
Connections in count	0.916369355331945
Connections out count	0.738347853121381
Bytes in count	0.828906532116767
Bytes out count	0.746395840638934
SYN-ONLY rate ratio	0.94681369012198
SYN-ACK rate ratio	0
Idle connection rate ratio	0.326352518624231
Half-open connection rate ratio	0.959963619166561
Packet Service discrete	0.28894159817916
Bytes Service discrete	0.265743563387272
Connection Service discrete	0.32071925304171
Packet Source port discrete	0.529763270923281
Bytes Source port discrete	0.470831403442902
Connection Source port discrete	0.52703826577568
Connection Source port orderedContinuous	0.453893438890411
Packet TTL discrete	0.5663269667398
Packet TTL orderedContinuous	0.625404058890487
InterPacket delta sortedContinuous	0.8804841513897972
Packet sec orderedContinuous	0.982606608342225
Packet min orderedContinuous	0.808898715948499
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.942425387371624
Bytes min orderedContinuous	0.774652098340554
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Continued on next page	

Table L.14 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.53498016377919
Packets in last w secs orderedContinuous	0.334619462160356
Priv packets time rate sortedContinuous	0.8384409985735419
Unpriv packets time rate sortedContinuous	0.8162418627269441
Connections time rate sortedContinuous	0.9569540410660404
Priv connections connection time rate sortedContinuous	0.6412441661406778
Unpriv connections connection time rate sortedContinuous	0.8386160240154669
Priv packets priv connection time rate sortedContinuous	0.7504880130577411
Unpriv packets unpriv connection time rate sortedContinuous	0.8760836465749871
SYNs connection time rate sortedContinuous	0.6444435718310096
RSTs connection time rate sortedContinuous	0.8261651376743373
FINs connection time rate sortedContinuous	0.8376752742746662
PSH connection time rate sortedContinuous	0.8467574990135538
Establishment errors connection time rate sortedContinuous	0.6627624954131978
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9569365631403456
Ave duration over last w secs sortedContinuous	0.8455604406228518
Priv packets packet rate sortedContinuous	0.8023423925203704
Unpriv packets packet rate sortedContinuous	0.8577644156338255
InterConnection delta sortedContinuous	0.9120624062525477
Connection sec orderedContinuous	0.975005600704244
Connection min orderedContinuous	0.718607029992827
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.8517306951653841
Connection Priv connections rate sortedContinuous	0.9032944970741652
Connection Unpriv connections rate sortedContinuous	0.8552095645769055
Connection Priv packet rate sortedContinuous	0.7160366813437988
Connection Unpriv packet rate sortedContinuous	0.8176560176893076
Continued on next page	

Table L.14 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.8710400115093781
Connection RSTs rate sortedContinuous	0.9932063260805509
Connection FINs rate sortedContinuous	0.9642532710862937
Connection PSH rate sortedContinuous	0.6413706213282384
Connection Establishment errors rate sortedContinuous	0.938679168124675
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9987743489610287
Ave duration over last m connections sortedContinuous	0.8030407309235441
Number of packets orderedContinuous	0.305149518557652
Number of packets in orderedContinuous	0.267436860210997
Number of packets out orderedContinuous	0.264403985272038
Duration sortedContinuous	0.9931481893186151
Number control packets rate sortedContinuous	0.9765964402288541
Number data packets rate sortedContinuous	0.9739263408369013
Number bytes transferred orderedContinuous	0.160032215638779
Number bytes transferred in orderedContinuous	0.146556516560344
Number bytes transferred out orderedContinuous	0.250193796730314
Number data bytes transferred orderedContinuous	0.140824408803922
Number data bytes transferred in orderedContinuous	0.126943558592925
Number data bytes transferred out orderedContinuous	0.239964044199884
Fragmented packets rate sortedContinuous	0.9922624994346938
Bad fragment rate sortedContinuous	0.9922624994346938
Max Src Window orderedContinuous	0.303591541979891
Max Dst Window orderedContinuous	0.144311253862924
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9914345167512043
Wrong resend rate sortedContinuous	0.9921434254679152
Duplicate ACK rate sortedContinuous	0.9999163074862953
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9997417566149476
Window exceeded rate sortedContinuous	0.9989451599834861
Continued on next page	

Table L.14 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9998274516138947
Number connection errors orderedContinuous	0.151855976806827
Number reset connection orderedContinuous	0.798032147879984
Number other errors orderedContinuous	0.791665800742012
Number disconnection errors orderedContinuous	0.33395462469716
Packet Destination IP nonKeyedSortedContinuous	0.968488579555295
Bytes Destination IP nonKeyedSortedContinuous	0.958175625518085
Connection Destination IP nonKeyedSortedContinuous	0.968111700630788
Packet Source IP nonKeyedSortedContinuous	0.81637442319982
Bytes Source IP nonKeyedSortedContinuous	0.816120386284231
Connection Source IP nonKeyedSortedContinuous	0.830709011276345

Table L.15: Similarity values for individual metrics of basecase 3,
test pair 1

Metric	Similarity
Packets in count	0.695375722543353
Packets out count	0.7516666666666667
Connections in count	0.898911353032659
Connections out count	0.961194029850746
Bytes in count	0.556519038323161
Bytes out count	0.721088923682924
SYN-ONLY rate ratio	0.713513513513514
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.899763220205208
Packet Service discrete	0.0620012906761658
Bytes Service discrete	0.0512338109209805
Connection Service discrete	0.153033446176722
Continued on next page	

Table L.15 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.0293318086380542
Bytes Source port discrete	0.025254271836069
Connection Source port discrete	0.0303854220970277
Connection Source port orderedContinuous	0.040509883644281
Packet TTL discrete	0.371164144952748
Packet TTL orderedContinuous	0.542575359685463
InterPacket delta sortedContinuous	0.8374709759783872
Packet sec orderedContinuous	0.498776870096425
Packet min orderedContinuous	0.631011217584437
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.464133465703377
Bytes min orderedContinuous	0.558332205670792
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.277538859982586
Packets in last w secs orderedContinuous	0.441574846555352
Priv packets time rate sortedContinuous	0.9644879926602319
Unpriv packets time rate sortedContinuous	0.9345897599277329
Connections time rate sortedContinuous	0.8302204860075351
Priv connections connection time rate sortedContinuous	0.9204486487330006
Unpriv connections connection time rate sortedContinuous	0.8166227640972144
Priv packets priv connection time rate sortedContinuous	0.8713126792953161
Unpriv packets unpriv connection time rate sortedContinuous	0.7524521528444222
SYNs connection time rate sortedContinuous	0.8064267917766771
RSTs connection time rate sortedContinuous	0.8151317885885868
FINs connection time rate sortedContinuous	0.6825890201069289
PSH connection time rate sortedContinuous	0.8608340775996166
Establishment errors connection time rate sortedContinuous	0.7282257155299992
Continued on next page	

Table L.15 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.2185862837881718
Priv packets packet rate sortedContinuous	0.96433383721778
Unpriv packets packet rate sortedContinuous	0.9570944432582523
InterConnection delta sortedContinuous	0.7674580152709915
Connection sec orderedContinuous	0.598286554044233
Connection min orderedContinuous	0.805240336413409
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.9001816701676863
Connection Priv connections rate sortedContinuous	0.9275390875485737
Connection Unpriv connections rate sortedContinuous	0.7587745426378711
Connection Priv packet rate sortedContinuous	0.8774726037295238
Connection Unpriv packet rate sortedContinuous	0.7851813088115534
Connection SYNs rate sortedContinuous	0.7102133884695683
Connection RSTs rate sortedContinuous	0.6743234343501229
Connection FINs rate sortedContinuous	0.8911877135924625
Connection PSH rate sortedContinuous	0.9045253368407417
Connection Establishment errors rate sortedContinuous	0.0641179854665423
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.138798847884184
Number of packets orderedContinuous	0.254098009336064
Number of packets in orderedContinuous	0.236195883803621
Number of packets out orderedContinuous	0.257672755476353
Duration sortedContinuous	0.8476418041422207
Number control packets rate sortedContinuous	0.9239049202500341
Number data packets rate sortedContinuous	0.929776416964816
Number bytes transferred orderedContinuous	0.398038412161522
Continued on next page	

Table L.15 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.381596531189178
Number bytes transferred out orderedContinuous	0.354976983581266
Number data bytes transferred orderedContinuous	0.17870234964335
Number data bytes transferred in orderedContinuous	0.151968878920876
Number data bytes transferred out orderedContinuous	0.250060310587734
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.555985324947589
Max Dst Window orderedContinuous	0.538929146537842
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9616737153659683
Wrong resend rate sortedContinuous	0.966068696982339
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9803921568627451
Hole rate sortedContinuous	0.9803921568627451
Number connection errors orderedContinuous	0.422480026612962
Number reset connection orderedContinuous	0.825495228556881
Number other errors orderedContinuous	0.798718633208756
Number disconnection errors orderedContinuous	0.798718633208756
Packet Destination IP nonKeyedSortedContinuous	0.867726138548677
Bytes Destination IP nonKeyedSortedContinuous	0.857799300928475
Connection Destination IP nonKeyedSortedContinuous	0.899573614069367
Packet Source IP nonKeyedSortedContinuous	0.923115150718138
Bytes Source IP nonKeyedSortedContinuous	0.860358143072572
Connection Source IP nonKeyedSortedContinuous	0.961058305284978

Table L.16: Similarity values for individual metrics of basecase 3,
test pair 2

Metric	Similarity
Packets in count	0.489184202091256
Packets out count	0.573665571479708
Connections in count	0.36968167836035
Connections out count	0.413545213772229
Bytes in count	0.106515135672464
Bytes out count	0.563732166834215
SYN-ONLY rate ratio	0.177603545253111
SYN-ACK rate ratio	0.609713282621417
Idle connection rate ratio	0.824693312227938
Half-open connection rate ratio	0.665882647137627
Packet Service discrete	0.0405426127603716
Bytes Service discrete	0.0206360497800969
Connection Service discrete	0.0629837343800997
Packet Source port discrete	0.0384764711166345
Bytes Source port discrete	0.0174248978728076
Connection Source port discrete	0.0318755804637872
Connection Source port orderedContinuous	0.0433964310255691
Packet TTL discrete	0.379708722040589
Packet TTL orderedContinuous	0.490045668463446
InterPacket delta sortedContinuous	0.686940451519534
Packet sec orderedContinuous	0.821168262103386
Packet min orderedContinuous	0.587347004319367
Packet GmHour orderedContinuous	0.148656241245144
Packet LocHour orderedContinuous	0.194172542813439
Packet weekday orderedContinuous	0.353191909810907
Bytes sec orderedContinuous	0.842691256417019
Bytes min orderedContinuous	0.417821692158971
Bytes GmHour orderedContinuous	0.145232537834284
Bytes LocHour orderedContinuous	0.182047979491192
Continued on next page	

Table L.16 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	0.347124606580712
Packet size orderedContinuous	0.153973636411557
Packets in last w secs orderedContinuous	0.609202603948227
Priv packets time rate sortedContinuous	0.9250509051005188
Unpriv packets time rate sortedContinuous	0.9182923451865248
Connections time rate sortedContinuous	0.7685039466940199
Priv connections connection time rate sortedContinuous	0.9749597941121817
Unpriv connections connection time rate sortedContinuous	0.9169617519535661
Priv packets priv connection time rate sortedContinuous	0.6809911428734532
Unpriv packets unpriv connection time rate sortedContinuous	0.8193701276848648
SYNs connection time rate sortedContinuous	0.9148369889555152
RSTs connection time rate sortedContinuous	0.7641452563729459
FINs connection time rate sortedContinuous	0.9029105827346336
PSH connection time rate sortedContinuous	0.5457666347072912
Establishment errors connection time rate sortedContinuous	0.5168022160121628
Other errors connection time rate sortedContinuous	0.9125365244934701
Disconnection errors connection time rate sortedContinuous	0.9440613532869546
Ave duration over last w secs sortedContinuous	0.6558339408812694
Priv packets packet rate sortedContinuous	0.9154454740075442
Unpriv packets packet rate sortedContinuous	0.9099178526283126
InterConnection delta sortedContinuous	0.5056207026054294
Connection sec orderedContinuous	0.745990824143095
Connection min orderedContinuous	0.590295649411598
Connection GmHour orderedContinuous	0.165385793045328
Connection LocHour orderedContinuous	0.218269328583634
Connection weekday orderedContinuous	0.388131706533653
Connection packet rate sortedContinuous	0.8398226140633255
Connection Priv connections rate sortedContinuous	0.9695742500169893
Connection Unpriv connections rate sortedContinuous	0.9002881538805044
Connection Priv packet rate sortedContinuous	0.746389299823021
Connection Unpriv packet rate sortedContinuous	0.6209310034058815
Continued on next page	

Table L.16 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.6143168364489892
Connection RSTs rate sortedContinuous	0.82480592788253
Connection FINs rate sortedContinuous	0.7253961914242143
Connection PSH rate sortedContinuous	0.7067957057732594
Connection Establishment errors rate sortedContinuous	0.7260820570897831
Connection Other errors rate sortedContinuous	0.9777308872767857
Connection Disconnection errors rate sortedContinuous	0.9846226283482143
Ave duration over last m connections sortedContinuous	0.5203847578321424
Number of packets orderedContinuous	0.199026303414005
Number of packets in orderedContinuous	0.194246378998342
Number of packets out orderedContinuous	0.199023462547195
Duration sortedContinuous	0.9477485457754672
Number control packets rate sortedContinuous	0.9597290342807476
Number data packets rate sortedContinuous	0.9691475771852417
Number bytes transferred orderedContinuous	0.12127037274929
Number bytes transferred in orderedContinuous	0.128601217935012
Number bytes transferred out orderedContinuous	0.122489912641393
Number data bytes transferred orderedContinuous	0.0954997646695661
Number data bytes transferred in orderedContinuous	0.073488923991584
Number data bytes transferred out orderedContinuous	0.103883029633753
Fragmented packets rate sortedContinuous	0.7236312751443065
Bad fragment rate sortedContinuous	0.7236312751443065
Max Src Window orderedContinuous	0.190681175720623
Max Dst Window orderedContinuous	0.175232403494077
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.7101126367677219
Wrong resend rate sortedContinuous	0.9033393329393462
Duplicate ACK rate sortedContinuous	0.9987073861436626
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9948816378758797
Window exceeded rate sortedContinuous	0.9550512277919712
Continued on next page	

Table L.16 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9914898623261448
Number connection errors orderedContinuous	0.16656047986822
Number reset connection orderedContinuous	0.58349183131551
Number other errors orderedContinuous	0.213387233833556
Number disconnection errors orderedContinuous	0.500213737032271
Packet Destination IP nonKeyedSortedContinuous	0.87616087162109
Bytes Destination IP nonKeyedSortedContinuous	0.856517856654307
Connection Destination IP nonKeyedSortedContinuous	0.916112975477011
Packet Source IP nonKeyedSortedContinuous	0.474380145187241
Bytes Source IP nonKeyedSortedContinuous	0.402391263700001
Connection Source IP nonKeyedSortedContinuous	0.471575408373548

Table L.17: Similarity values for individual metrics of basecase 3,
test pair 3

Metric	Similarity
Packets in count	0.178652830338143
Packets out count	0.242503658827133
Connections in count	0.990656597977729
Connections out count	0.975260804769001
Bytes in count	0.283114299234705
Bytes out count	0.208191337293614
SYN-ONLY rate ratio	0.689545725054553
SYN-ACK rate ratio	1
Idle connection rate ratio	0.354260089686098
Half-open connection rate ratio	0.815455952976586
Packet Service discrete	0.0274025108371442
Bytes Service discrete	0.031000552415156
Connection Service discrete	0.0571044318401538
Continued on next page	

Table L.17 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.0286256986468484
Bytes Source port discrete	0.0323692249085684
Connection Source port discrete	0.0914900325431399
Connection Source port orderedContinuous	0.143829926296193
Packet TTL discrete	0.194151292488804
Packet TTL orderedContinuous	0.249006127693697
InterPacket delta sortedContinuous	0.5136891906577901
Packet sec orderedContinuous	0.77261433588209
Packet min orderedContinuous	0.813830837785955
Packet GmHour orderedContinuous	0.183603792402662
Packet LocHour orderedContinuous	0.163203371024588
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.800950693688069
Bytes min orderedContinuous	0.710246315539969
Bytes GmHour orderedContinuous	0.188406379147196
Bytes LocHour orderedContinuous	0.16747233701973
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.368229108304034
Packets in last w secs orderedContinuous	0.120365524128052
Priv packets time rate sortedContinuous	0.8727393913702422
Unpriv packets time rate sortedContinuous	0.8978810931973931
Connections time rate sortedContinuous	0.182989453345278
Priv connections connection time rate sortedContinuous	0.9795436912872817
Unpriv connections connection time rate sortedContinuous	0.6580054886040094
Priv packets priv connection time rate sortedContinuous	0.1813201062738394
Unpriv packets unpriv connection time rate sortedContinuous	0.3367921987173444
SYNs connection time rate sortedContinuous	0.709275374560081
RSTs connection time rate sortedContinuous	0.7012027846371862
FINs connection time rate sortedContinuous	0.7732005256936658
PSH connection time rate sortedContinuous	0.06794700592746689
Establishment errors connection time rate sortedContinuous	0.8922491505785765
Continued on next page	

Table L.17 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9851460494121214
Ave duration over last w secs sortedContinuous	0.6360212787699916
Priv packets packet rate sortedContinuous	0.8153402332728031
Unpriv packets packet rate sortedContinuous	0.8498446781897953
InterConnection delta sortedContinuous	0.6475276214238268
Connection sec orderedContinuous	0.517646209543971
Connection min orderedContinuous	0.936803269132424
Connection GmHour orderedContinuous	0.191471597835253
Connection LocHour orderedContinuous	0.170196975853558
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.1245614393097329
Connection Priv connections rate sortedContinuous	0.992295109584174
Connection Unpriv connections rate sortedContinuous	0.6266934565197627
Connection Priv packet rate sortedContinuous	0.1487291359208191
Connection Unpriv packet rate sortedContinuous	0.5671005916196268
Connection SYNs rate sortedContinuous	0.7490689790623754
Connection RSTs rate sortedContinuous	0.7830728998808473
Connection FINs rate sortedContinuous	0.7442124129162124
Connection PSH rate sortedContinuous	0.07715158357532345
Connection Establishment errors rate sortedContinuous	0.9634620339569457
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9945768839679132
Ave duration over last m connections sortedContinuous	0.6078578078615196
Number of packets orderedContinuous	0.362858960035652
Number of packets in orderedContinuous	0.347011860081166
Number of packets out orderedContinuous	0.35863292149677
Duration sortedContinuous	0.9511711021516694
Number control packets rate sortedContinuous	0.9683387927537205
Number data packets rate sortedContinuous	0.9860323473087077
Number bytes transferred orderedContinuous	0.11189480941941
Continued on next page	

Table L.17 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.107509211569274
Number bytes transferred out orderedContinuous	0.232384756156758
Number data bytes transferred orderedContinuous	0.0741766383032685
Number data bytes transferred in orderedContinuous	0.0808684704496303
Number data bytes transferred out orderedContinuous	0.176137963036232
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.476717980113806
Max Dst Window orderedContinuous	0.133325724351696
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9869698507521122
Wrong resend rate sortedContinuous	0.9953076234982021
Duplicate ACK rate sortedContinuous	0.998849252013809
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9931648261815238
Hole rate sortedContinuous	0.9976985040276179
Number connection errors orderedContinuous	0.89267537749516
Number reset connection orderedContinuous	0.882231315193327
Number other errors orderedContinuous	0.891952739301036
Number disconnection errors orderedContinuous	0.912684432788612
Packet Destination IP nonKeyedSortedContinuous	0.892824527481349
Bytes Destination IP nonKeyedSortedContinuous	0.879499945229068
Connection Destination IP nonKeyedSortedContinuous	0.955600133217001
Packet Source IP nonKeyedSortedContinuous	0.854776021412259
Bytes Source IP nonKeyedSortedContinuous	0.816275968282066
Connection Source IP nonKeyedSortedContinuous	0.81609715946724

Table L.18: Similarity values for individual metrics of basecase 4,
train pair 1

Metric	Similarity
Packets in count	0.712121212121212
Packets out count	0.728476821192053
Connections in count	0.637992831541219
Connections out count	0.646502835538752
Bytes in count	0.693548138203478
Bytes out count	0.789774953205937
SYN-ONLY rate ratio	0.69387755102041
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.69387755102041
Packet Service discrete	0.016080343233284
Bytes Service discrete	0.0160197982290148
Connection Service discrete	0.018926384663469
Packet Source port discrete	0.00858924425121448
Bytes Source port discrete	0.00868550443155265
Connection Source port discrete	0.00432526431651797
Connection Source port orderedContinuous	0.013138571714154
Packet TTL discrete	0.226457493561992
Packet TTL orderedContinuous	0.440951701763717
InterPacket delta sortedContinuous	0.8313283284271861
Packet sec orderedContinuous	0.629698595205682
Packet min orderedContinuous	0.291950921752115
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.510996332527546
Bytes min orderedContinuous	0.311108399152316
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Continued on next page	

Table L.18 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.334599563621567
Packets in last w secs orderedContinuous	0.566271698896123
Priv packets time rate sortedContinuous	0.970625816766538
Unpriv packets time rate sortedContinuous	0.9695713924916893
Connections time rate sortedContinuous	0.9044953689998184
Priv connections connection time rate sortedContinuous	0.9810475243799026
Unpriv connections connection time rate sortedContinuous	0.9965834428383706
Priv packets priv connection time rate sortedContinuous	0.8727304513071302
Unpriv packets unpriv connection time rate sortedContinuous	0.996933858957512
SYNs connection time rate sortedContinuous	0.9022169793554889
RSTs connection time rate sortedContinuous	0.9248588349555399
FINs connection time rate sortedContinuous	0.7005824025266468
PSH connection time rate sortedContinuous	0.9054293066559611
Establishment errors connection time rate sortedContinuous	0.9224704336399474
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.6101345833127282
Priv packets packet rate sortedContinuous	0.9679953160471049
Unpriv packets packet rate sortedContinuous	0.9743862531550745
InterConnection delta sortedContinuous	0.7966289841918145
Connection sec orderedContinuous	0.535463162350397
Connection min orderedContinuous	0.309959736294445
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.8984468647669971
Connection Priv connections rate sortedContinuous	0.9851012868620398
Connection Unpriv connections rate sortedContinuous	0.5827229627026483
Connection Priv packet rate sortedContinuous	0.8908638958250362
Connection Unpriv packet rate sortedContinuous	0.6236867239732569
Continued on next page	

Table L.18 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.8114687163771114
Connection RSTs rate sortedContinuous	0.5979213109392755
Connection FINs rate sortedContinuous	0.794847246964673
Connection PSH rate sortedContinuous	0.7595938631711238
Connection Establishment errors rate sortedContinuous	0.8051575931232091
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.7632084597834276
Number of packets orderedContinuous	0.316273121673269
Number of packets in orderedContinuous	0.458986819326831
Number of packets out orderedContinuous	0.378199780766829
Duration sortedContinuous	0.9887027594713702
Number control packets rate sortedContinuous	0.993933456436605
Number data packets rate sortedContinuous	0.9937695058669342
Number bytes transferred orderedContinuous	0.230780810076228
Number bytes transferred in orderedContinuous	0.250158688911584
Number bytes transferred out orderedContinuous	0.188817582918391
Number data bytes transferred orderedContinuous	0.0436594400197036
Number data bytes transferred in orderedContinuous	0.0809292125992644
Number data bytes transferred out orderedContinuous	0.146490294287027
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.118840579710145
Max Dst Window orderedContinuous	0.246849518161601
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.8848740070720855
Wrong resend rate sortedContinuous	0.9411764705882353
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9794973223098022
Continued on next page	

Table L.18 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.843184328590881
Number reset connection orderedContinuous	0.91779774868741
Number other errors orderedContinuous	0.941890866155014
Number disconnection errors orderedContinuous	0.941890866155014
Packet Destination IP nonKeyedSortedContinuous	0.814822902655871
Bytes Destination IP nonKeyedSortedContinuous	0.820534345817371
Connection Destination IP nonKeyedSortedContinuous	0.736470218063422
Packet Source IP nonKeyedSortedContinuous	0.854994923251986
Bytes Source IP nonKeyedSortedContinuous	0.696521468556583
Connection Source IP nonKeyedSortedContinuous	0.91340293651009

Table L.19: Similarity values for individual metrics of basecase 4,
train pair 2

Metric	Similarity
Packets in count	0.0359577335494834
Packets out count	0.134327926677653
Connections in count	0.914149443561208
Connections out count	0.648090815273478
Bytes in count	0.00316042304202879
Bytes out count	0.0559730283357699
SYN-ONLY rate ratio	0.67212770426381
SYN-ACK rate ratio	1
Idle connection rate ratio	0.37313432835821
Half-open connection rate ratio	0.693787448754336
Packet Service discrete	0.00226318718174551
Bytes Service discrete	0.000204337927507751
Connection Service discrete	0.0441285105186106
Continued on next page	

Table L.19 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.00217247327696567
Bytes Source port discrete	0.000205653363016461
Connection Source port discrete	0.0266685489736241
Connection Source port orderedContinuous	0.0430120534148642
Packet TTL discrete	0.0629640669154507
Packet TTL orderedContinuous	0.159409483555671
InterPacket delta sortedContinuous	0.1396947558361044
Packet sec orderedContinuous	0.813514247574042
Packet min orderedContinuous	0.300080307257004
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.741333216883704
Bytes min orderedContinuous	0.21068064338761
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.0339110737749185
Packets in last w secs orderedContinuous	0.00630289823694239
Priv packets time rate sortedContinuous	0.06275558842403987
Unpriv packets time rate sortedContinuous	0.8077620185192742
Connections time rate sortedContinuous	0.06930160588488557
Priv connections connection time rate sortedContinuous	0.7681637182713377
Unpriv connections connection time rate sortedContinuous	0.5922456615602544
Priv packets priv connection time rate sortedContinuous	0.7963918675743287
Unpriv packets unpriv connection time rate sortedContinuous	0.07805303290315604
SYNs connection time rate sortedContinuous	0.5451550235493953
RSTs connection time rate sortedContinuous	0.3962579573949182
FINs connection time rate sortedContinuous	0.8396081760427671
PSH connection time rate sortedContinuous	0.1122375257281266
Establishment errors connection time rate sortedContinuous	0.425581565357884
Continued on next page	

Table L.19 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.01181680266255553
Priv packets packet rate sortedContinuous	0.04612458073894486
Unpriv packets packet rate sortedContinuous	0.7892762927339392
InterConnection delta sortedContinuous	0.8202636440005429
Connection sec orderedContinuous	0.76485529231651
Connection min orderedContinuous	0.739675107669756
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.3593271848977652
Connection Priv connections rate sortedContinuous	0.801191942818783
Connection Unpriv connections rate sortedContinuous	0.4476032265363842
Connection Priv packet rate sortedContinuous	0.6922378902812653
Connection Unpriv packet rate sortedContinuous	0.737056755182904
Connection SYNs rate sortedContinuous	0.6478021806132464
Connection RSTs rate sortedContinuous	0.8639593433638911
Connection FINs rate sortedContinuous	0.7903434032706708
Connection PSH rate sortedContinuous	0.5790462896013236
Connection Establishment errors rate sortedContinuous	0.88262660376646
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.04628055186402083
Number of packets orderedContinuous	0.176053944901231
Number of packets in orderedContinuous	0.194365603680046
Number of packets out orderedContinuous	0.174757294564298
Duration sortedContinuous	0.6847946174386179
Number control packets rate sortedContinuous	0.8927614705188175
Number data packets rate sortedContinuous	0.9003848135807132
Number bytes transferred orderedContinuous	0.220568114606567
Continued on next page	

Table L.19 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.139099291566985
Number bytes transferred out orderedContinuous	0.196465330907771
Number data bytes transferred orderedContinuous	0.0612137379475787
Number data bytes transferred in orderedContinuous	0.0624442211820275
Number data bytes transferred out orderedContinuous	0.0857007499384201
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.100982076040233
Max Dst Window orderedContinuous	0.126978538796514
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9357798165137615
Wrong resend rate sortedContinuous	0.9388379204892966
Duplicate ACK rate sortedContinuous	0.9541284403669725
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9877675840978594
Hole rate sortedContinuous	0.9541284403669725
Number connection errors orderedContinuous	0.902087033841137
Number reset connection orderedContinuous	0.901586308430925
Number other errors orderedContinuous	0.901255090501233
Number disconnection errors orderedContinuous	0.901255090501233
Packet Destination IP nonKeyedSortedContinuous	0.621462169388022
Bytes Destination IP nonKeyedSortedContinuous	0.437087176495021
Connection Destination IP nonKeyedSortedContinuous	0.714184448214853
Packet Source IP nonKeyedSortedContinuous	0.683027217893155
Bytes Source IP nonKeyedSortedContinuous	0.435311308914089
Connection Source IP nonKeyedSortedContinuous	0.480250041479384

Table L.20: Similarity values for individual metrics of basecase 4,
train pair 3

Metric	Similarity
Packets in count	0.0034863451481697
Packets out count	0.00241254523522316
Connections in count	0.00619834710743805
Connections out count	0.6666666666666667
Bytes in count	0.00844727456874728
Bytes out count	0.00629811554127535
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	2.21196919738909e-05
Bytes Service discrete	7.60012484689277e-05
Connection Service discrete	0.000641215943444474
Packet Source port discrete	2.10953380200297e-05
Bytes Source port discrete	5.62889495266107e-05
Connection Source port discrete	2.99326142022771e-05
Connection Source port orderedContinuous	8.14987068871841e-05
Packet TTL discrete	0.149676160056783
Packet TTL orderedContinuous	0.24218007050005
InterPacket delta sortedContinuous	0.2054438294454655
Packet sec orderedContinuous	0.0613678669700471
Packet min orderedContinuous	0
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	0.25
Bytes sec orderedContinuous	0.0446508537996196
Bytes min orderedContinuous	0
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Continued on next page	

Table L.20 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	0.25
Packet size orderedContinuous	0.0452632546148214
Packets in last w secs orderedContinuous	0.000269019939239265
Priv packets time rate sortedContinuous	0.9536738052099228
Unpriv packets time rate sortedContinuous	0.9448869507754744
Connections time rate sortedContinuous	0.588567769132785
Priv connections connection time rate sortedContinuous	0.9
Unpriv connections connection time rate sortedContinuous	0.9857142857142857
Priv packets priv connection time rate sortedContinuous	0.7381285057681858
Unpriv packets unpriv connection time rate sortedContinuous	1
SYNs connection time rate sortedContinuous	0.5389539000425124
RSTs connection time rate sortedContinuous	0.62
FINs connection time rate sortedContinuous	0.1
PSH connection time rate sortedContinuous	0.255731188830955
Establishment errors connection time rate sortedContinuous	0.4
Other errors connection time rate sortedContinuous	0.9
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.1008960739327695
Priv packets packet rate sortedContinuous	0.9087106470586691
Unpriv packets packet rate sortedContinuous	0.841191068129081
InterConnection delta sortedContinuous	0.4392620859359723
Connection sec orderedContinuous	0.0427961052564981
Connection min orderedContinuous	0
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	0.25
Connection packet rate sortedContinuous	0.4553001296983153
Connection Priv connections rate sortedContinuous	0.5809523809523811
Connection Unpriv connections rate sortedContinuous	0.07627118644067796
Connection Priv packet rate sortedContinuous	0.4100921730979536
Connection Unpriv packet rate sortedContinuous	0.25
Continued on next page	

Table L.20 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.05102040816326531
Connection RSTs rate sortedContinuous	0.5
Connection FINs rate sortedContinuous	0.25
Connection PSH rate sortedContinuous	0.25
Connection Establishment errors rate sortedContinuous	0.5
Connection Other errors rate sortedContinuous	0.75
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.25
Number of packets orderedContinuous	0.22553739990023
Number of packets in orderedContinuous	0.286787586049424
Number of packets out orderedContinuous	0.276319112998745
Duration sortedContinuous	0.2503218054514588
Number control packets rate sortedContinuous	0.4318181818181817
Number data packets rate sortedContinuous	0.2999999999999999
Number bytes transferred orderedContinuous	0.0046089993651281
Number bytes transferred in orderedContinuous	0.00110414596802998
Number bytes transferred out orderedContinuous	0.00367195995083394
Number data bytes transferred orderedContinuous	0.0155210747683404
Number data bytes transferred in orderedContinuous	0.0155210747683404
Number data bytes transferred out orderedContinuous	0.191226985886123
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.00130491518051328
Max Dst Window orderedContinuous	0.00125104253544622
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	1
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	1
Continued on next page	

Table L.20 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.263482254686713
Number reset connection orderedContinuous	0.408699783284284
Number other errors orderedContinuous	0.590292146834669
Number disconnection errors orderedContinuous	0.52115332794395
Packet Destination IP nonKeyedSortedContinuous	0.430314601145513
Bytes Destination IP nonKeyedSortedContinuous	0.249538799124237
Connection Destination IP nonKeyedSortedContinuous	0.336432506887052
Packet Source IP nonKeyedSortedContinuous	0.224793204449519
Bytes Source IP nonKeyedSortedContinuous	0.130728425400402
Connection Source IP nonKeyedSortedContinuous	0.52519408594175

Table L.21: Similarity values for individual metrics of basecase 4,
test pair 1

Metric	Similarity
Packets in count	0.874680306905371
Packets out count	0.844933285250631
Connections in count	0.932061978545888
Connections out count	0.857142857142857
Bytes in count	0.981821304371895
Bytes out count	0.673191278493558
SYN-ONLY rate ratio	0.630278063851699
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.928881179531655
Packet Service discrete	0.0229375224138485
Bytes Service discrete	0.0233563817675041
Connection Service discrete	0.074160301506316
Continued on next page	

Table L.21 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.0144084937838227
Bytes Source port discrete	0.0142828218771593
Connection Source port discrete	0.011499647447893
Connection Source port orderedContinuous	0.0278868654128634
Packet TTL discrete	0.400905833834151
Packet TTL orderedContinuous	0.503591373913971
InterPacket delta sortedContinuous	0.827386750979465
Packet sec orderedContinuous	0.441803290569133
Packet min orderedContinuous	0.567287105627302
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.410420827952026
Bytes min orderedContinuous	0.490853754730517
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.451584731861404
Packets in last w secs orderedContinuous	0.451839782205633
Priv packets time rate sortedContinuous	0.9737597458908517
Unpriv packets time rate sortedContinuous	0.9716176638655307
Connections time rate sortedContinuous	0.9521294790044308
Priv connections connection time rate sortedContinuous	0.9825569765201997
Unpriv connections connection time rate sortedContinuous	0.7352263382146168
Priv packets priv connection time rate sortedContinuous	0.9672616794085801
Unpriv packets unpriv connection time rate sortedContinuous	0.691704874483372
SYNs connection time rate sortedContinuous	0.9469065596395228
RSTs connection time rate sortedContinuous	0.6725052919202776
FINs connection time rate sortedContinuous	0.7396355582446203
PSH connection time rate sortedContinuous	0.8995525109103323
Establishment errors connection time rate sortedContinuous	0.7285612455872058
Continued on next page	

Table L.21 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.6028667341123959
Priv packets packet rate sortedContinuous	0.9850181044189063
Unpriv packets packet rate sortedContinuous	0.9871543119872232
InterConnection delta sortedContinuous	0.8968430818965019
Connection sec orderedContinuous	0.564945278871858
Connection min orderedContinuous	0.69530911571221
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.9797129869417598
Connection Priv connections rate sortedContinuous	0.9856124677463182
Connection Unpriv connections rate sortedContinuous	0.8722529732981142
Connection Priv packet rate sortedContinuous	0.9657842281935997
Connection Unpriv packet rate sortedContinuous	0.9404869860327053
Connection SYNs rate sortedContinuous	0.8338739675584338
Connection RSTs rate sortedContinuous	0.7440766103708124
Connection FINs rate sortedContinuous	0.9324442920226001
Connection PSH rate sortedContinuous	0.9573852488912654
Connection Establishment errors rate sortedContinuous	0.4611380334643018
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.6532974771327491
Number of packets orderedContinuous	0.399987517091299
Number of packets in orderedContinuous	0.441373563279197
Number of packets out orderedContinuous	0.551454314274229
Duration sortedContinuous	0.8916372185882712
Number control packets rate sortedContinuous	0.9053273984765065
Number data packets rate sortedContinuous	0.9205148942456796
Number bytes transferred orderedContinuous	0.126417110569711
Continued on next page	

Table L.21 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.148566386360375
Number bytes transferred out orderedContinuous	0.111349197283006
Number data bytes transferred orderedContinuous	0.0499705591298362
Number data bytes transferred in orderedContinuous	0.098771261084768
Number data bytes transferred out orderedContinuous	0.0933512959574417
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.409326147858258
Max Dst Window orderedContinuous	0.0600142489942164
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9467775225627174
Wrong resend rate sortedContinuous	0.9842105263157894
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9947368421052631
Hole rate sortedContinuous	0.9894736842105263
Number connection errors orderedContinuous	0.295676844104401
Number reset connection orderedContinuous	0.866132355094657
Number other errors orderedContinuous	0.870075211958842
Number disconnection errors orderedContinuous	0.870075211958842
Packet Destination IP nonKeyedSortedContinuous	0.881385044700415
Bytes Destination IP nonKeyedSortedContinuous	0.862641035043753
Connection Destination IP nonKeyedSortedContinuous	0.941643379481234
Packet Source IP nonKeyedSortedContinuous	0.908050604966385
Bytes Source IP nonKeyedSortedContinuous	0.879587634590099
Connection Source IP nonKeyedSortedContinuous	0.944183819585783

Table L.22: Similarity values for individual metrics of basecase 4,
test pair 2

Metric	Similarity
Packets in count	0.161062118422328
Packets out count	0.00186950909772243
Connections in count	0.155821917808219
Connections out count	0.00233696705494402
Bytes in count	0.161180939213184
Bytes out count	0.0019348282808147
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	0.00113742260856501
Bytes Service discrete	0.00113372677186931
Connection Service discrete	0.000498778650021568
Packet Source port discrete	2.29330103969848e-05
Bytes Source port discrete	2.37125633146771e-05
Connection Source port discrete	1.67325662035914e-05
Connection Source port orderedContinuous	5.82615594222878e-05
Packet TTL discrete	0.0272173121191963
Packet TTL orderedContinuous	0.0382927671815205
InterPacket delta sortedContinuous	0.3267838250559983
Packet sec orderedContinuous	0.836006925127956
Packet min orderedContinuous	0.792905108755863
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.79117221992782
Bytes min orderedContinuous	0.76587041918378
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Continued on next page	

Table L.22 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.0370558052606922
Packets in last w secs orderedContinuous	8.89618941011654e-05
Priv packets time rate sortedContinuous	0.2409823868022093
Unpriv packets time rate sortedContinuous	0.01934770880233858
Connections time rate sortedContinuous	0.8423817590085722
Priv connections connection time rate sortedContinuous	0.1621168462061412
Unpriv connections connection time rate sortedContinuous	0.01361815566392173
Priv packets priv connection time rate sortedContinuous	0.2725144806998653
Unpriv packets unpriv connection time rate sortedContinuous	0.8875880174985873
SYNs connection time rate sortedContinuous	0.1808199410507827
RSTs connection time rate sortedContinuous	0.9875
FINs connection time rate sortedContinuous	0.6589435179698716
PSH connection time rate sortedContinuous	0.4112195518301094
Establishment errors connection time rate sortedContinuous	0.22265625
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.1578125
Priv packets packet rate sortedContinuous	0.4107538514479587
Unpriv packets packet rate sortedContinuous	0.01997197219286371
InterConnection delta sortedContinuous	0.2982030104071018
Connection sec orderedContinuous	0.859644543065021
Connection min orderedContinuous	0.78059405273473
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.9409429979653657
Connection Priv connections rate sortedContinuous	0.4236149791303295
Connection Unpriv connections rate sortedContinuous	0.0222110920612028
Connection Priv packet rate sortedContinuous	0.4731999273393254
Connection Unpriv packet rate sortedContinuous	0.07130095417803416
Continued on next page	

Table L.22 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.7960681082595059
Connection RSTs rate sortedContinuous	0.9990503323836657
Connection FINs rate sortedContinuous	0.8795504906616018
Connection PSH rate sortedContinuous	0.6858587799186695
Connection Establishment errors rate sortedContinuous	0.8366571699905033
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.8338081671415005
Number of packets orderedContinuous	0.116086131972203
Number of packets in orderedContinuous	0.168512647721272
Number of packets out orderedContinuous	0.240609766237654
Duration sortedContinuous	0.9791073124406457
Number control packets rate sortedContinuous	0.9791073124406457
Number data packets rate sortedContinuous	0.9791073124406457
Number bytes transferred orderedContinuous	0.0412635767640847
Number bytes transferred in orderedContinuous	0.043340637408332
Number bytes transferred out orderedContinuous	0.0295964035152512
Number data bytes transferred orderedContinuous	0.0712813092554554
Number data bytes transferred in orderedContinuous	0.0965230402023476
Number data bytes transferred out orderedContinuous	0.0915384501551705
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0
Max Dst Window orderedContinuous	0
Urgent rate sortedContinuous	0
Resend rate sortedContinuous	0
Wrong resend rate sortedContinuous	0
Duplicate ACK rate sortedContinuous	0
Wrong ACK sortedContinuous	0
Wrong data packet size rate sortedContinuous	0
Window exceeded rate sortedContinuous	0
Continued on next page	

Table L.22 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0
Number connection errors orderedContinuous	0
Number reset connection orderedContinuous	0
Number other errors orderedContinuous	0
Number disconnection errors orderedContinuous	0
Packet Destination IP nonKeyedSortedContinuous	0.29049943472085
Bytes Destination IP nonKeyedSortedContinuous	0.274915986488801
Connection Destination IP nonKeyedSortedContinuous	0.320997369136537
Packet Source IP nonKeyedSortedContinuous	0.245255328556215
Bytes Source IP nonKeyedSortedContinuous	0.257195048811535
Connection Source IP nonKeyedSortedContinuous	0.290535917480807

Table L.23: Similarity values for individual metrics of basecase 5,
train pair 1

Metric	Similarity
Packets in count	0.0757377379393696
Packets out count	0.0650008441668074
Connections in count	0.862619808306709
Connections out count	0.824742268041237
Bytes in count	0.126581957621068
Bytes out count	0.00784586328388093
SYN-ONLY rate ratio	0.521739130434782
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.521739130434782
Packet Service discrete	0.00167722563621722
Bytes Service discrete	0.000593164696071417
Connection Service discrete	0.0351288606190054
Continued on next page	

Table L.23 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.00083677745423656
Bytes Source port discrete	0.00028361206191768
Connection Source port discrete	0.00531179925433097
Connection Source port orderedContinuous	0.0146074479494102
Packet TTL discrete	0.0859215459591998
Packet TTL orderedContinuous	0.171785913226053
InterPacket delta sortedContinuous	0.5879389249069953
Packet sec orderedContinuous	0.748810803891598
Packet min orderedContinuous	0.177896402773009
Packet GmHour orderedContinuous	0
Packet LocHour orderedContinuous	0
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.703625016344549
Bytes min orderedContinuous	0.135264481402846
Bytes GmHour orderedContinuous	0
Bytes LocHour orderedContinuous	0
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.0757126504150433
Packets in last w secs orderedContinuous	0.0377412866451722
Priv packets time rate sortedContinuous	0.9503173301606048
Unpriv packets time rate sortedContinuous	0.9386111858581273
Connections time rate sortedContinuous	0.03651754187115273
Priv connections connection time rate sortedContinuous	0.9850197109067017
Unpriv connections connection time rate sortedContinuous	0.9947437582128777
Priv packets priv connection time rate sortedContinuous	0.04381787007975084
Unpriv packets unpriv connection time rate sortedContinuous	0.9947437582128777
SYNs connection time rate sortedContinuous	0.05364915109680061
RSTs connection time rate sortedContinuous	0.6917612391988213
FINs connection time rate sortedContinuous	0.4369324106130491
PSH connection time rate sortedContinuous	0.03296226917429108
Establishment errors connection time rate sortedContinuous	0.9986859395532195
Continued on next page	

Table L.23 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.1613055704710301
Priv packets packet rate sortedContinuous	0.9720797779515943
Unpriv packets packet rate sortedContinuous	0.970995310856617
InterConnection delta sortedContinuous	0.8255689148301291
Connection sec orderedContinuous	0.44202296667399
Connection min orderedContinuous	0.202782800404959
Connection GmHour orderedContinuous	0
Connection LocHour orderedContinuous	0
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.3600768877795187
Connection Priv connections rate sortedContinuous	0.9794098398940195
Connection Unpriv connections rate sortedContinuous	0.5529411764705883
Connection Priv packet rate sortedContinuous	0.3614678727388783
Connection Unpriv packet rate sortedContinuous	0.5529411764705883
Connection SYNs rate sortedContinuous	0.6610002051989928
Connection RSTs rate sortedContinuous	0.6900161232380431
Connection FINs rate sortedContinuous	0.772854392491465
Connection PSH rate sortedContinuous	0.2396661868054529
Connection Establishment errors rate sortedContinuous	0.7803921568627451
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.763743556813806
Number of packets orderedContinuous	0.372153424642389
Number of packets in orderedContinuous	0.56539019886473
Number of packets out orderedContinuous	0.607290148083438
Duration sortedContinuous	0.9661424355263678
Number control packets rate sortedContinuous	0.9717181339629962
Number data packets rate sortedContinuous	0.9731987285508956
Number bytes transferred orderedContinuous	0.295985807629117
Continued on next page	

Table L.23 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.271937361443423
Number bytes transferred out orderedContinuous	0.244999356289938
Number data bytes transferred orderedContinuous	0.0726993013069001
Number data bytes transferred in orderedContinuous	0.143115286497329
Number data bytes transferred out orderedContinuous	0.186157299404821
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.28698752228164
Max Dst Window orderedContinuous	0.194835680751174
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.8778280542986425
Wrong resend rate sortedContinuous	0.9411764705882353
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9388235294117647
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.823560652920056
Number reset connection orderedContinuous	0.856165970185815
Number other errors orderedContinuous	0.816614862955637
Number disconnection errors orderedContinuous	0.816614862955637
Packet Destination IP nonKeyedSortedContinuous	0.677414671301338
Bytes Destination IP nonKeyedSortedContinuous	0.713567387938556
Connection Destination IP nonKeyedSortedContinuous	0.83654697085602
Packet Source IP nonKeyedSortedContinuous	0.700058489048563
Bytes Source IP nonKeyedSortedContinuous	0.612786622109964
Connection Source IP nonKeyedSortedContinuous	0.948954006261465

Table L.24: Similarity values for individual metrics of basecase 5,
train pair 2

Metric	Similarity
Packets in count	0.0466909293802863
Packets out count	0.0727756892230577
Connections in count	0.10907711277625
Connections out count	0.0796549974632167
Bytes in count	0.00622179174054582
Bytes out count	0.0784907351119259
SYN-ONLY rate ratio	0.409646391896287
SYN-ACK rate ratio	0
Idle connection rate ratio	0.33030303030303
Half-open connection rate ratio	0.388028169014083
Packet Service discrete	0.000768624072361653
Bytes Service discrete	0.000385957724490575
Connection Service discrete	0.000641722852092567
Packet Source port discrete	0.00058345341322471
Bytes Source port discrete	0.000337431529408507
Connection Source port discrete	0.000858619637765093
Connection Source port orderedContinuous	0.00411512157628553
Packet TTL discrete	0.0402264882444846
Packet TTL orderedContinuous	0.100165706201064
InterPacket delta sortedContinuous	0.3921493796347462
Packet sec orderedContinuous	0.822053358516569
Packet min orderedContinuous	0.809488502109225
Packet GmHour orderedContinuous	0
Packet LocHour orderedContinuous	0
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.739514737016179
Bytes min orderedContinuous	0.719952174925534
Bytes GmHour orderedContinuous	0
Bytes LocHour orderedContinuous	0
Continued on next page	

Table L.24 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.0611936063973772
Packets in last w secs orderedContinuous	0.000513624204744539
Priv packets time rate sortedContinuous	0.837770613918752
Unpriv packets time rate sortedContinuous	0.924129222272453
Connections time rate sortedContinuous	0.2885078006761988
Priv connections connection time rate sortedContinuous	0.7803921850092324
Unpriv connections connection time rate sortedContinuous	0.7289775987262642
Priv packets priv connection time rate sortedContinuous	0.3660629030092936
Unpriv packets unpriv connection time rate sortedContinuous	0.1634177628381455
SYNs connection time rate sortedContinuous	0.4741999711315442
RSTs connection time rate sortedContinuous	0.558430502867916
FINs connection time rate sortedContinuous	0.3602768149090529
PSH connection time rate sortedContinuous	0.4243870705575137
Establishment errors connection time rate sortedContinuous	0.4938079135835955
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.01665090326893244
Priv packets packet rate sortedContinuous	0.8251902225980676
Unpriv packets packet rate sortedContinuous	0.9387223504151265
InterConnection delta sortedContinuous	0.1416829137268065
Connection sec orderedContinuous	0.800890029428921
Connection min orderedContinuous	0.584534253449915
Connection GmHour orderedContinuous	0
Connection LocHour orderedContinuous	0
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.07565104814382931
Connection Priv connections rate sortedContinuous	0.7893853674942257
Connection Unpriv connections rate sortedContinuous	0.6023892703099474
Connection Priv packet rate sortedContinuous	0.4962237853179179
Connection Unpriv packet rate sortedContinuous	0.02688205766444933

Continued on next page

Table L.24 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.7206725859210508
Connection RSTs rate sortedContinuous	0.3685176764835866
Connection FINs rate sortedContinuous	0.5071279165355237
Connection PSH rate sortedContinuous	0.09141378709799211
Connection Establishment errors rate sortedContinuous	0.4178863022391697
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.08203675667365739
Number of packets orderedContinuous	0.0809279139575212
Number of packets in orderedContinuous	0.0877084545226818
Number of packets out orderedContinuous	0.11598868939552
Duration sortedContinuous	0.7664430232486472
Number control packets rate sortedContinuous	0.8052176894280609
Number data packets rate sortedContinuous	0.8054770327482694
Number bytes transferred orderedContinuous	0.0196258015684661
Number bytes transferred in orderedContinuous	0.013067916045028
Number bytes transferred out orderedContinuous	0.0173814107279321
Number data bytes transferred orderedContinuous	0.0034362177714726
Number data bytes transferred in orderedContinuous	0.00413178954914646
Number data bytes transferred out orderedContinuous	0.00656868171837111
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0852695400126203
Max Dst Window orderedContinuous	0.00990909465197841
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9694189602446484
Wrong resend rate sortedContinuous	0.9908256880733946
Duplicate ACK rate sortedContinuous	0.9969418960244648
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9847094801223242
Window exceeded rate sortedContinuous	0.9969418960244648
Continued on next page	

Table L.24 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9969418960244648
Number connection errors orderedContinuous	0.309224884018493
Number reset connection orderedContinuous	0.672023105566318
Number other errors orderedContinuous	0.677668331419559
Number disconnection errors orderedContinuous	0.677668331419559
Packet Destination IP nonKeyedSortedContinuous	0.703746633011774
Bytes Destination IP nonKeyedSortedContinuous	0.60403101258871
Connection Destination IP nonKeyedSortedContinuous	0.775487179130651
Packet Source IP nonKeyedSortedContinuous	0.668250363767712
Bytes Source IP nonKeyedSortedContinuous	0.518636218396662
Connection Source IP nonKeyedSortedContinuous	0.48815103721952

Table L.25: Similarity values for individual metrics of basecase 5,
train pair 3

Metric	Similarity
Packets in count	0.909090909090909
Packets out count	1
Connections in count	0.8
Connections out count	1
Bytes in count	0.652920962199313
Bytes out count	1
SYN-ONLY rate ratio	1
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	1
Packet Service discrete	0.270676691729323
Bytes Service discrete	0.229407665505226
Connection Service discrete	0.342857142857143
Continued on next page	

Table L.25 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.270676691729323
Bytes Source port discrete	0.229407665505226
Connection Source port discrete	0.142857142857143
Connection Source port orderedContinuous	0.142857142857143
Packet TTL discrete	0.417484388938448
Packet TTL orderedContinuous	0.493182107811903
InterPacket delta sortedContinuous	0.4968707179083702
Packet sec orderedContinuous	0
Packet min orderedContinuous	0
Packet GmHour orderedContinuous	0
Packet LocHour orderedContinuous	0
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0
Bytes min orderedContinuous	0
Bytes GmHour orderedContinuous	0
Bytes LocHour orderedContinuous	0
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.863777089783281
Packets in last w secs orderedContinuous	0.940733714098514
Priv packets time rate sortedContinuous	0.8810381942912063
Unpriv packets time rate sortedContinuous	0.9514387768735474
Connections time rate sortedContinuous	0.9713501763806023
Priv connections connection time rate sortedContinuous	0.8888888888888888
Unpriv connections connection time rate sortedContinuous	0.8888888888888888
Priv packets priv connection time rate sortedContinuous	0.8655325181640971
Unpriv packets unpriv connection time rate sortedContinuous	0.8888888888888888
SYNs connection time rate sortedContinuous	0.9470899470899471
RSTs connection time rate sortedContinuous	1
FINs connection time rate sortedContinuous	1
PSH connection time rate sortedContinuous	0.9528619528619528
Establishment errors connection time rate sortedContinuous	1
Continued on next page	

Table L.25 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.7925870460434702
Priv packets packet rate sortedContinuous	0.7711812351582791
Unpriv packets packet rate sortedContinuous	0.7488410258279677
InterConnection delta sortedContinuous	0.3493460298683424
Connection sec orderedContinuous	0
Connection min orderedContinuous	0
Connection GmHour orderedContinuous	0
Connection LocHour orderedContinuous	0
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	1
Connection Priv connections rate sortedContinuous	0.5313283208020051
Connection Unpriv connections rate sortedContinuous	0
Connection Priv packet rate sortedContinuous	0.6666666666666666
Connection Unpriv packet rate sortedContinuous	0
Connection SYNs rate sortedContinuous	0.9523809523809526
Connection RSTs rate sortedContinuous	1
Connection FINs rate sortedContinuous	1
Connection PSH rate sortedContinuous	1
Connection Establishment errors rate sortedContinuous	1
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	1
Number of packets orderedContinuous	0.899159663865546
Number of packets in orderedContinuous	0.937441643323996
Number of packets out orderedContinuous	0.91880458722564
Duration sortedContinuous	0.919480370577666
Number control packets rate sortedContinuous	1
Number data packets rate sortedContinuous	1
Number bytes transferred orderedContinuous	0.589285714285714
Continued on next page	

Table L.25 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.696428571428572
Number bytes transferred out orderedContinuous	0.841558441558442
Number data bytes transferred orderedContinuous	0.899159663865546
Number data bytes transferred in orderedContinuous	0.899159663865546
Number data bytes transferred out orderedContinuous	0.919327731092437
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0
Max Dst Window orderedContinuous	1
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	1
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	1
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.857142857142857
Number reset connection orderedContinuous	0.857142857142857
Number other errors orderedContinuous	0.857142857142857
Number disconnection errors orderedContinuous	0.857142857142857
Packet Destination IP nonKeyedSortedContinuous	0.954545454545455
Bytes Destination IP nonKeyedSortedContinuous	0.822768589680745
Connection Destination IP nonKeyedSortedContinuous	0.9
Packet Source IP nonKeyedSortedContinuous	0.7
Bytes Source IP nonKeyedSortedContinuous	1
Connection Source IP nonKeyedSortedContinuous	1

Table L.26: Similarity values for individual metrics of basecase 5,
test pair 1

Metric	Similarity
Packets in count	0.936126088759851
Packets out count	0.78000520697735
Connections in count	0.950335570469799
Connections out count	0.610526315789474
Bytes in count	0.898126360602645
Bytes out count	0.591232352813076
SYN-ONLY rate ratio	0.195386702849389
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.653414001728607
Packet Service discrete	0.0364160161240795
Bytes Service discrete	0.0385931822865411
Connection Service discrete	0.0875605397075319
Packet Source port discrete	0.0220678837907739
Bytes Source port discrete	0.0238270138922789
Connection Source port discrete	0.0174118334838539
Connection Source port orderedContinuous	0.0348051035786494
Packet TTL discrete	0.302168417356362
Packet TTL orderedContinuous	0.471993677371262
InterPacket delta sortedContinuous	0.819273703207945
Packet sec orderedContinuous	0.554603072023266
Packet min orderedContinuous	0.529725072037268
Packet GmHour orderedContinuous	0
Packet LocHour orderedContinuous	0
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.418550132665769
Bytes min orderedContinuous	0.447877233571031
Bytes GmHour orderedContinuous	0
Bytes LocHour orderedContinuous	0
Continued on next page	

Table L.26 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.26286709055023
Packets in last w secs orderedContinuous	0.617530472277718
Priv packets time rate sortedContinuous	0.8830011209208934
Unpriv packets time rate sortedContinuous	0.9205171561713622
Connections time rate sortedContinuous	0.731314352407896
Priv connections connection time rate sortedContinuous	0.843574022636735
Unpriv connections connection time rate sortedContinuous	0.7030512328659479
Priv packets priv connection time rate sortedContinuous	0.7409483437108223
Unpriv packets unpriv connection time rate sortedContinuous	0.6668443458429239
SYNs connection time rate sortedContinuous	0.7688308956714301
RSTs connection time rate sortedContinuous	0.9538608200983347
FINs connection time rate sortedContinuous	0.8409887287114732
PSH connection time rate sortedContinuous	0.7679212256608512
Establishment errors connection time rate sortedContinuous	0.714435675159348
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.5130967603471565
Priv packets packet rate sortedContinuous	0.9114185581734382
Unpriv packets packet rate sortedContinuous	0.9294173776578554
InterConnection delta sortedContinuous	0.5642476664098092
Connection sec orderedContinuous	0.648703389746604
Connection min orderedContinuous	0.670158503938327
Connection GmHour orderedContinuous	0
Connection LocHour orderedContinuous	0
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.8185259043491877
Connection Priv connections rate sortedContinuous	0.8313804124598101
Connection Unpriv connections rate sortedContinuous	0.3563793824547516
Connection Priv packet rate sortedContinuous	0.8586266626166282
Connection Unpriv packet rate sortedContinuous	0.601230059236433
Continued on next page	

Table L.26 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.7795850065033315
Connection RSTs rate sortedContinuous	0.3639094817377188
Connection FINs rate sortedContinuous	0.6839421266556013
Connection PSH rate sortedContinuous	0.8291551991564195
Connection Establishment errors rate sortedContinuous	0.01760715026772634
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.393905309641958
Number of packets orderedContinuous	0.203837936265326
Number of packets in orderedContinuous	0.221933318001879
Number of packets out orderedContinuous	0.326406958346275
Duration sortedContinuous	0.8879397558858482
Number control packets rate sortedContinuous	0.916631512363899
Number data packets rate sortedContinuous	0.9093076948600658
Number bytes transferred orderedContinuous	0.140014552178501
Number bytes transferred in orderedContinuous	0.133061678505356
Number bytes transferred out orderedContinuous	0.106110312616686
Number data bytes transferred orderedContinuous	0.0569633204450905
Number data bytes transferred in orderedContinuous	0.0452756091275762
Number data bytes transferred out orderedContinuous	0.0922160531335902
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.450460577673692
Max Dst Window orderedContinuous	0.0941042607079356
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.5945638170880166
Wrong resend rate sortedContinuous	0.966666666666666667
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.988888888888888889
Continued on next page	

Table L.26 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.284506548481549
Number reset connection orderedContinuous	0.853206821617578
Number other errors orderedContinuous	0.827695004382121
Number disconnection errors orderedContinuous	0.827695004382121
Packet Destination IP nonKeyedSortedContinuous	0.817328920947462
Bytes Destination IP nonKeyedSortedContinuous	0.860007163603076
Connection Destination IP nonKeyedSortedContinuous	0.903220592654967
Packet Source IP nonKeyedSortedContinuous	0.752766111396394
Bytes Source IP nonKeyedSortedContinuous	0.745896244600298
Connection Source IP nonKeyedSortedContinuous	0.901548673906146

Table L.27: Similarity values for individual metrics of basecase 5,
test pair 2

Metric	Similarity
Packets in count	0.0616527080784643
Packets out count	0.0560763126607128
Connections in count	0.845544706696394
Connections out count	0.727522802953525
Bytes in count	0.030049270889219
Bytes out count	0.0189987278518662
SYN-ONLY rate ratio	0.660847157352231
SYN-ACK rate ratio	0.333116672083198
Idle connection rate ratio	0
Half-open connection rate ratio	0.524410817442869
Packet Service discrete	0.0293089314381806
Bytes Service discrete	0.0288982487274638
Connection Service discrete	0.062526540940753
Continued on next page	

Table L.27 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.0286046347318449
Bytes Source port discrete	0.0230257976393892
Connection Source port discrete	0.101522746998537
Connection Source port orderedContinuous	0.184168074823319
Packet TTL discrete	0.159087555757618
Packet TTL orderedContinuous	0.216026235321381
InterPacket delta sortedContinuous	0.7573437226083994
Packet sec orderedContinuous	0.95153358432184
Packet min orderedContinuous	0.807410941015705
Packet GmHour orderedContinuous	0
Packet LocHour orderedContinuous	0
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.911711588681182
Bytes min orderedContinuous	0.797800495073313
Bytes GmHour orderedContinuous	0
Bytes LocHour orderedContinuous	0
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.410195367805454
Packets in last w secs orderedContinuous	0.000852426185602094
Priv packets time rate sortedContinuous	0.9235435960176754
Unpriv packets time rate sortedContinuous	0.9552737355867617
Connections time rate sortedContinuous	0.1337854194397798
Priv connections connection time rate sortedContinuous	0.8530111446647811
Unpriv connections connection time rate sortedContinuous	0.9414734383926312
Priv packets priv connection time rate sortedContinuous	0.2084613513894619
Unpriv packets unpriv connection time rate sortedContinuous	0.1101958325640918
SYNs connection time rate sortedContinuous	0.6757651046039175
RSTs connection time rate sortedContinuous	0.2138276462377324
FINs connection time rate sortedContinuous	0.719664012104569
PSH connection time rate sortedContinuous	0.1310684269804997
Establishment errors connection time rate sortedContinuous	0.4362397535446977
Continued on next page	

Table L.27 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9883564387012097
Ave duration over last w secs sortedContinuous	0.5649237954668294
Priv packets packet rate sortedContinuous	0.910980396759208
Unpriv packets packet rate sortedContinuous	0.9508195405102366
InterConnection delta sortedContinuous	0.7588931355058136
Connection sec orderedContinuous	0.977783341973748
Connection min orderedContinuous	0.956060653002777
Connection GmHour orderedContinuous	0
Connection LocHour orderedContinuous	0
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.1023726600623141
Connection Priv connections rate sortedContinuous	0.6166688096861923
Connection Unpriv connections rate sortedContinuous	0.6738968143252256
Connection Priv packet rate sortedContinuous	0.2275808719905041
Connection Unpriv packet rate sortedContinuous	0.1964861455840643
Connection SYNs rate sortedContinuous	0.4985115510629522
Connection RSTs rate sortedContinuous	0.465492236508515
Connection FINs rate sortedContinuous	0.5597781756153306
Connection PSH rate sortedContinuous	0.1152001304256934
Connection Establishment errors rate sortedContinuous	0.7414531504314232
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.999619916381604
Ave duration over last m connections sortedContinuous	0.1955140924612857
Number of packets orderedContinuous	0.0997675525953046
Number of packets in orderedContinuous	0.0811025637745771
Number of packets out orderedContinuous	0.0767753212635409
Duration sortedContinuous	0.8551686825168538
Number control packets rate sortedContinuous	0.8675772411047554
Number data packets rate sortedContinuous	0.9291616065903072
Number bytes transferred orderedContinuous	0.11652257571818
Continued on next page	

Table L.27 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.133798595088298
Number bytes transferred out orderedContinuous	0.118639471563419
Number data bytes transferred orderedContinuous	0.0913006442109322
Number data bytes transferred in orderedContinuous	0.122037648794525
Number data bytes transferred out orderedContinuous	0.0888569545476557
Fragmented packets rate sortedContinuous	0.9998086215970527
Bad fragment rate sortedContinuous	0.9998086215970527
Max Src Window orderedContinuous	0.117503100295137
Max Dst Window orderedContinuous	0.0277040723834158
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9829633519946594
Wrong resend rate sortedContinuous	0.9863307424283034
Duplicate ACK rate sortedContinuous	0.99812382739212
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9983467389890959
Hole rate sortedContinuous	0.99906191369606
Number connection errors orderedContinuous	0.415477570943251
Number reset connection orderedContinuous	0.469329698739727
Number other errors orderedContinuous	0.48696717001402
Number disconnection errors orderedContinuous	0.586426225179638
Packet Destination IP nonKeyedSortedContinuous	0.756625557814846
Bytes Destination IP nonKeyedSortedContinuous	0.697964145919948
Connection Destination IP nonKeyedSortedContinuous	0.906958477685831
Packet Source IP nonKeyedSortedContinuous	0.735719264966773
Bytes Source IP nonKeyedSortedContinuous	0.598566965129157
Connection Source IP nonKeyedSortedContinuous	0.921746871131478

Table L.28: Similarity values for individual metrics of basecase 5,
test pair 3

Metric	Similarity
Packets in count	0
Packets out count	0
Connections in count	0
Connections out count	0
Bytes in count	0
Bytes out count	0
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	0
Bytes Service discrete	0
Connection Service discrete	0
Packet Source port discrete	0
Bytes Source port discrete	0
Connection Source port discrete	0
Connection Source port orderedContinuous	0
Packet TTL discrete	0
Packet TTL orderedContinuous	0
InterPacket delta sortedContinuous	0
Packet sec orderedContinuous	0
Packet min orderedContinuous	0
Packet GmHour orderedContinuous	0
Packet LocHour orderedContinuous	0
Packet weekday orderedContinuous	0
Bytes sec orderedContinuous	0
Bytes min orderedContinuous	0
Bytes GmHour orderedContinuous	0
Bytes LocHour orderedContinuous	0
Continued on next page	

Table L.28 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	0
Packet size orderedContinuous	0
Packets in last w secs orderedContinuous	0
Priv packets time rate sortedContinuous	0
Unpriv packets time rate sortedContinuous	0
Connections time rate sortedContinuous	0
Priv connections connection time rate sortedContinuous	0
Unpriv connections connection time rate sortedContinuous	0
Priv packets priv connection time rate sortedContinuous	0
Unpriv packets unpriv connection time rate sortedContinuous	0
SYNs connection time rate sortedContinuous	0
RSTs connection time rate sortedContinuous	0
FINs connection time rate sortedContinuous	0
PSH connection time rate sortedContinuous	0
Establishment errors connection time rate sortedContinuous	0
Other errors connection time rate sortedContinuous	0
Disconnection errors connection time rate sortedContinuous	0
Ave duration over last w secs sortedContinuous	0
Priv packets packet rate sortedContinuous	0
Unpriv packets packet rate sortedContinuous	0
InterConnection delta sortedContinuous	0
Connection sec orderedContinuous	0
Connection min orderedContinuous	0
Connection GmHour orderedContinuous	0
Connection LocHour orderedContinuous	0
Connection weekday orderedContinuous	0
Connection packet rate sortedContinuous	0
Connection Priv connections rate sortedContinuous	0
Connection Unpriv connections rate sortedContinuous	0
Connection Priv packet rate sortedContinuous	0
Connection Unpriv packet rate sortedContinuous	0
Continued on next page	

Table L.28 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0
Connection RSTs rate sortedContinuous	0
Connection FINs rate sortedContinuous	0
Connection PSH rate sortedContinuous	0
Connection Establishment errors rate sortedContinuous	0
Connection Other errors rate sortedContinuous	0
Connection Disconnection errors rate sortedContinuous	0
Ave duration over last m connections sortedContinuous	0
Number of packets orderedContinuous	0
Number of packets in orderedContinuous	0
Number of packets out orderedContinuous	0
Duration sortedContinuous	0
Number control packets rate sortedContinuous	0
Number data packets rate sortedContinuous	0
Number bytes transferred orderedContinuous	0
Number bytes transferred in orderedContinuous	0
Number bytes transferred out orderedContinuous	0
Number data bytes transferred orderedContinuous	0
Number data bytes transferred in orderedContinuous	0
Number data bytes transferred out orderedContinuous	0
Fragmented packets rate sortedContinuous	0
Bad fragment rate sortedContinuous	0
Max Src Window orderedContinuous	0
Max Dst Window orderedContinuous	0
Urgent rate sortedContinuous	0
Resend rate sortedContinuous	0
Wrong resend rate sortedContinuous	0
Duplicate ACK rate sortedContinuous	0
Wrong ACK sortedContinuous	0
Wrong data packet size rate sortedContinuous	0
Window exceeded rate sortedContinuous	0
Continued on next page	

Table L.28 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0
Number connection errors orderedContinuous	0
Number reset connection orderedContinuous	0
Number other errors orderedContinuous	0
Number disconnection errors orderedContinuous	0
Packet Destination IP nonKeyedSortedContinuous	0
Bytes Destination IP nonKeyedSortedContinuous	0
Connection Destination IP nonKeyedSortedContinuous	0
Packet Source IP nonKeyedSortedContinuous	0
Bytes Source IP nonKeyedSortedContinuous	0
Connection Source IP nonKeyedSortedContinuous	0

Table L.29: Similarity values for individual metrics of basecase 6,
train pair 1

Metric	Similarity
Packets in count	0.90631636907103
Connections in count	0.788990825688073
Bytes in count	0.787132127080225
SYN-ONLY rate ratio	1
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	1
Packet Service discrete	0.137363346108798
Bytes Service discrete	0.110765016499258
Connection Service discrete	0.719709899826464
Packet Source port discrete	0.054799070461651
Bytes Source port discrete	0.032819548487308
Connection Source port discrete	0.0146109412164458
Continued on next page	

Table L.29 – continued from previous page

Metric	Similarity
Connection Source port orderedContinuous	0.0402546339636772
Packet TTL discrete	0.224613553649068
Packet TTL orderedContinuous	0.321823799454987
InterPacket delta sortedContinuous	0.8266296753708596
Packet sec orderedContinuous	0.979581150159518
Packet min orderedContinuous	0.779504309550039
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.966811774810175
Bytes min orderedContinuous	0.535252026567194
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.465549196483057
Packets in last w secs orderedContinuous	0.509857348737735
Priv packets time rate sortedContinuous	0.9683417445325576
Unpriv packets time rate sortedContinuous	0.998315335948792
Connections time rate sortedContinuous	0.9183438371910787
Priv connections connection time rate sortedContinuous	0.8578466698586186
Unpriv connections connection time rate sortedContinuous	0.927898186259552
Priv packets priv connection time rate sortedContinuous	0.9543805977057137
Unpriv packets unpriv connection time rate sortedContinuous	0.8510727243474169
SYNs connection time rate sortedContinuous	0.9660150378545629
RSTs connection time rate sortedContinuous	0.9841544584546694
FINs connection time rate sortedContinuous	0.9009962453770718
PSH connection time rate sortedContinuous	0.9185081470041598
Establishment errors connection time rate sortedContinuous	0.9706111148490688
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.9309259800857967
Continued on next page	

Table L.29 – continued from previous page

Metric	Similarity
Priv packets packet rate sortedContinuous	0.9897075726053695
Unpriv packets packet rate sortedContinuous	0.998893737642232
InterConnection delta sortedContinuous	0.6791959086081011
Connection sec orderedContinuous	0.636622664160275
Connection min orderedContinuous	0.288476805310329
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.4846256784215141
Connection Priv connections rate sortedContinuous	0.3441640932897428
Connection Unpriv connections rate sortedContinuous	0.8525026342165178
Connection Priv packet rate sortedContinuous	0.310293320866303
Connection Unpriv packet rate sortedContinuous	0.3902661383250141
Connection SYNs rate sortedContinuous	0.346882713454618
Connection RSTs rate sortedContinuous	0.4030665458084557
Connection FINs rate sortedContinuous	0.8323006835023689
Connection PSH rate sortedContinuous	0.4665733089110459
Connection Establishment errors rate sortedContinuous	0.6375854718074594
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.4896653800963193
Number of packets orderedContinuous	0.692972895039359
Number of packets in orderedContinuous	0.692972895039359
Number of packets out orderedContinuous	0.9676532575827
Duration sortedContinuous	0.8936546676517724
Number control packets rate sortedContinuous	0.9816083916083916
Number data packets rate sortedContinuous	0.9720238095238096
Number bytes transferred orderedContinuous	0.318728088855625
Number bytes transferred in orderedContinuous	0.318728088855625
Number bytes transferred out orderedContinuous	0.9676532575827
Number data bytes transferred orderedContinuous	0.426585138411993
Continued on next page	

Table L.29 – continued from previous page

Metric	Similarity
Number data bytes transferred in orderedContinuous	0.426585138411993
Number data bytes transferred out orderedContinuous	0.9676532575827
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.574712643678161
Max Dst Window orderedContinuous	0.808510638297872
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	1
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9796860572483841
Window exceeded rate sortedContinuous	1
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.67596276003819
Number reset connection orderedContinuous	0.979658258746949
Number other errors orderedContinuous	0.979658258746949
Number disconnection errors orderedContinuous	0.979658258746949
Packet Destination IP nonKeyedSortedContinuous	0.643210697708143
Bytes Destination IP nonKeyedSortedContinuous	0.617156285872168
Connection Destination IP nonKeyedSortedContinuous	0.820880193971327
Packet Source IP nonKeyedSortedContinuous	0.645555933744711
Bytes Source IP nonKeyedSortedContinuous	0.613951094406
Connection Source IP nonKeyedSortedContinuous	0.787561517429939

Table L.30: Similarity values for individual metrics of basecase 6,
train pair 2

Metric	Similarity
Packets in count	0.484158515426588
Packets out count	0.0772530625828043
Connections in count	0.42259671436142
Connections out count	0.0161600377423339
Bytes in count	0.472509894698613
Bytes out count	0.0454058147103945
SYN-ONLY rate ratio	0.183913088503246
SYN-ACK rate ratio	0
Idle connection rate ratio	0.723270374730438
Half-open connection rate ratio	0.257377005809221
Packet Service discrete	0.033272522618771
Bytes Service discrete	0.0315262548945743
Connection Service discrete	0.0229478966034085
Packet Source port discrete	0.0969483606254083
Bytes Source port discrete	0.0739619605350072
Connection Source port discrete	0.0332317429159862
Connection Source port orderedContinuous	0.0477988632375203
Packet TTL discrete	0.193082308052283
Packet TTL orderedContinuous	0.259566680766666
InterPacket delta sortedContinuous	0.426520941879878
Packet sec orderedContinuous	0.946516595033156
Packet min orderedContinuous	0.826313989426595
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.916178145969094
Bytes min orderedContinuous	0.785501641013744
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Continued on next page	

Table L.30 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.402797985123131
Packets in last w secs orderedContinuous	0.00752369990594898
Priv packets time rate sortedContinuous	0.7573520360842652
Unpriv packets time rate sortedContinuous	0.2312192651935219
Connections time rate sortedContinuous	0.1336618862747552
Priv connections connection time rate sortedContinuous	0.2176150030961843
Unpriv connections connection time rate sortedContinuous	0.0276000993898892
Priv packets priv connection time rate sortedContinuous	0.5600041331965088
Unpriv packets unpriv connection time rate sortedContinuous	0.6971415734456659
SYNs connection time rate sortedContinuous	0.3166808166467798
RSTs connection time rate sortedContinuous	0.02327368590685995
FINs connection time rate sortedContinuous	0.03879383519500166
PSH connection time rate sortedContinuous	0.01902520268243156
Establishment errors connection time rate sortedContinuous	0.2810966467367508
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.925210235680514
Ave duration over last w secs sortedContinuous	0.07926124490773789
Priv packets packet rate sortedContinuous	0.7243067938351226
Unpriv packets packet rate sortedContinuous	0.1986125597341041
InterConnection delta sortedContinuous	0.4927596120002939
Connection sec orderedContinuous	0.959555051031847
Connection min orderedContinuous	0.525389589186077
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.3765133524112635
Connection Priv connections rate sortedContinuous	0.2175694477847886
Connection Unpriv connections rate sortedContinuous	0.1239887039693212
Connection Priv packet rate sortedContinuous	0.2491588347808103
Connection Unpriv packet rate sortedContinuous	0.06199078639143591
Continued on next page	

Table L.30 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.5590825791823814
Connection RSTs rate sortedContinuous	0.4640739030733323
Connection FINs rate sortedContinuous	0.3926789953898729
Connection PSH rate sortedContinuous	0.1364268875210743
Connection Establishment errors rate sortedContinuous	0.6490553120996899
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9997149047781959
Ave duration over last m connections sortedContinuous	0.1033445104073736
Number of packets orderedContinuous	0.0543458710979663
Number of packets in orderedContinuous	0.044302332156937
Number of packets out orderedContinuous	0.0488133076267396
Duration sortedContinuous	0.8293339629843768
Number control packets rate sortedContinuous	0.8729468934795949
Number data packets rate sortedContinuous	0.8963757795173284
Number bytes transferred orderedContinuous	0.0145752864487441
Number bytes transferred in orderedContinuous	0.0163389820186689
Number bytes transferred out orderedContinuous	0.0332479872118882
Number data bytes transferred orderedContinuous	0.00949218285635679
Number data bytes transferred in orderedContinuous	0.0107476273462939
Number data bytes transferred out orderedContinuous	0.0332980832060043
Fragmented packets rate sortedContinuous	0.9981020302524874
Bad fragment rate sortedContinuous	0.9981020302524874
Max Src Window orderedContinuous	0.138875633133926
Max Dst Window orderedContinuous	0.0764138879602211
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9737769988912896
Wrong resend rate sortedContinuous	0.9938577262989246
Duplicate ACK rate sortedContinuous	0.999829902913817
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9852718625013512
Window exceeded rate sortedContinuous	0.9988819380767116
Continued on next page	

Table L.30 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9990589500571351
Number connection errors orderedContinuous	0.0969516499770649
Number reset connection orderedContinuous	0.449784235296944
Number other errors orderedContinuous	0.503357591806641
Number disconnection errors orderedContinuous	0.282188611209478
Packet Destination IP nonKeyedSortedContinuous	0.108849099318525
Bytes Destination IP nonKeyedSortedContinuous	0.0968996586016105
Connection Destination IP nonKeyedSortedContinuous	0.460487188644732
Packet Source IP nonKeyedSortedContinuous	0.211423347426474
Bytes Source IP nonKeyedSortedContinuous	0.114704781011791
Connection Source IP nonKeyedSortedContinuous	0.115807737959454

Table L.31: Similarity values for individual metrics of basecase 6,
train pair 3

Metric	Similarity
Packets in count	0.589532502822588
Packets out count	0.557340363847213
Connections in count	0.405079752108097
Connections out count	0.345978062157221
Bytes in count	0.614101381092414
Bytes out count	0.294640961811881
SYN-ONLY rate ratio	0.670153023771072
SYN-ACK rate ratio	0
Idle connection rate ratio	0.27831378138362
Half-open connection rate ratio	0.99543316209283
Packet Service discrete	0.0438473500592365
Bytes Service discrete	0.0319712144705921
Connection Service discrete	0.0114239065828007
Continued on next page	

Table L.31 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.055910027111076
Bytes Source port discrete	0.0332050223460192
Connection Source port discrete	0.03259207491068
Connection Source port orderedContinuous	0.0449244971126019
Packet TTL discrete	0.15329933265094
Packet TTL orderedContinuous	0.226252784220525
InterPacket delta sortedContinuous	0.6193733421692793
Packet sec orderedContinuous	0.923164325380816
Packet min orderedContinuous	0.718511591885024
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.916006670371338
Bytes min orderedContinuous	0.650517797799018
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.529578585063589
Packets in last w secs orderedContinuous	0.1289993891225
Priv packets time rate sortedContinuous	0.7191977400167509
Unpriv packets time rate sortedContinuous	0.8777929108499568
Connections time rate sortedContinuous	0.9300939415591947
Priv connections connection time rate sortedContinuous	0.9623616619554299
Unpriv connections connection time rate sortedContinuous	0.6688514114827321
Priv packets priv connection time rate sortedContinuous	0.7606339830183082
Unpriv packets unpriv connection time rate sortedContinuous	0.7826862201090788
SYNs connection time rate sortedContinuous	0.8058212548368526
RSTs connection time rate sortedContinuous	0.5455155448652962
FINs connection time rate sortedContinuous	0.5892465888870392
PSH connection time rate sortedContinuous	0.888526358097625
Establishment errors connection time rate sortedContinuous	0.3092720963064147
Continued on next page	

Table L.31 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9981954018840005
Ave duration over last w secs sortedContinuous	0.8270246916506504
Priv packets packet rate sortedContinuous	0.7149914677414687
Unpriv packets packet rate sortedContinuous	0.888200966107731
InterConnection delta sortedContinuous	0.549974055262772
Connection sec orderedContinuous	0.865735611440111
Connection min orderedContinuous	0.566941700088078
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.1166799493877007
Connection Priv connections rate sortedContinuous	0.2152329507189796
Connection Unpriv connections rate sortedContinuous	0.223955953383047
Connection Priv packet rate sortedContinuous	0.1039118612794057
Connection Unpriv packet rate sortedContinuous	0.03707369759178933
Connection SYNs rate sortedContinuous	0.4995104602753716
Connection RSTs rate sortedContinuous	0.5481196588536593
Connection FINs rate sortedContinuous	0.6752272234871298
Connection PSH rate sortedContinuous	0.115049483912411
Connection Establishment errors rate sortedContinuous	0.4292743618883634
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9997719238225568
Ave duration over last m connections sortedContinuous	0.3020791501106615
Number of packets orderedContinuous	0.205799587741963
Number of packets in orderedContinuous	0.202055103750075
Number of packets out orderedContinuous	0.177714489748114
Duration sortedContinuous	0.8446515642915597
Number control packets rate sortedContinuous	0.861726874287818
Number data packets rate sortedContinuous	0.8446704824563057
Number bytes transferred orderedContinuous	0.077378708242515
Continued on next page	

Table L.31 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.0818199881437851
Number bytes transferred out orderedContinuous	0.143320710987441
Number data bytes transferred orderedContinuous	0.0635523586081043
Number data bytes transferred in orderedContinuous	0.0587555396169138
Number data bytes transferred out orderedContinuous	0.148179595227803
Fragmented packets rate sortedContinuous	0.9998274572956807
Bad fragment rate sortedContinuous	0.9998274572956807
Max Src Window orderedContinuous	0.0866799634733347
Max Dst Window orderedContinuous	0.0778824646081967
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9740062219495543
Wrong resend rate sortedContinuous	0.9874252764804611
Duplicate ACK rate sortedContinuous	0.999664653473399
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9853780388451066
Window exceeded rate sortedContinuous	0.9933070289608191
Hole rate sortedContinuous	0.9985469529577723
Number connection errors orderedContinuous	0.140364225922983
Number reset connection orderedContinuous	0.498915661471943
Number other errors orderedContinuous	0.526188999731198
Number disconnection errors orderedContinuous	0.439070187624349
Packet Destination IP nonKeyedSortedContinuous	0.900997276438465
Bytes Destination IP nonKeyedSortedContinuous	0.838598506707062
Connection Destination IP nonKeyedSortedContinuous	0.882403719002435
Packet Source IP nonKeyedSortedContinuous	0.815747585132114
Bytes Source IP nonKeyedSortedContinuous	0.761065546062504
Connection Source IP nonKeyedSortedContinuous	0.613377037388744

Table L.32: Similarity values for individual metrics of basemode 6,
test pair 1

Metric	Similarity
Packets in count	0.870473859991402
Connections in count	0.75
Bytes in count	0.816090472537047
SYN-ONLY rate ratio	1
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	1
Packet Service discrete	0.0752569681577955
Bytes Service discrete	0.0413389698135293
Connection Service discrete	0.137513513513514
Packet Source port discrete	0.0200964347821242
Bytes Source port discrete	0.0186580577677601
Connection Source port discrete	0.0171428571428571
Connection Source port orderedContinuous	0.0475175957634458
Packet TTL discrete	0.28632876105071
Packet TTL orderedContinuous	0.468471494097632
InterPacket delta sortedContinuous	0.7670855802543903
Packet sec orderedContinuous	0.889155445688551
Packet min orderedContinuous	0.580309884519194
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.829862232526378
Bytes min orderedContinuous	0.469241806285329
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.60836101320397
Packets in last w secs orderedContinuous	0.493260056735214
Continued on next page	

Table L.32 – continued from previous page

Metric	Similarity
Priv packets time rate sortedContinuous	0.9625195357808103
Unpriv packets time rate sortedContinuous	0.9997668188920744
Connections time rate sortedContinuous	0.803519058708515
Priv connections connection time rate sortedContinuous	0.9095497669489561
Unpriv connections connection time rate sortedContinuous	0.6394057108015259
Priv packets priv connection time rate sortedContinuous	0.9625783387321276
Unpriv packets unpriv connection time rate sortedContinuous	0.6797768677367769
SYNs connection time rate sortedContinuous	0.9268356184114293
RSTs connection time rate sortedContinuous	0.8948375131114923
FINs connection time rate sortedContinuous	0.8852757207546674
PSH connection time rate sortedContinuous	0.7807405997334493
Establishment errors connection time rate sortedContinuous	0.7669346533883592
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.7730525277662621
Priv packets packet rate sortedContinuous	0.9929848553917844
Unpriv packets packet rate sortedContinuous	0.9998236635332379
InterConnection delta sortedContinuous	0.5911727786539107
Connection sec orderedContinuous	0.750685506785723
Connection min orderedContinuous	0.722391199463323
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.7925737998003776
Connection Priv connections rate sortedContinuous	0.8516400569203457
Connection Unpriv connections rate sortedContinuous	0.6862116730211193
Connection Priv packet rate sortedContinuous	0.9116499120250601
Connection Unpriv packet rate sortedContinuous	0.3512283574530886
Connection SYNs rate sortedContinuous	0.9501748107873058
Connection RSTs rate sortedContinuous	0.630717033660916
Connection FINs rate sortedContinuous	0.9682050840942553
Continued on next page	

Table L.32 – continued from previous page

Metric	Similarity
Connection PSH rate sortedContinuous	0.4953429814966175
Connection Establishment errors rate sortedContinuous	0.6101545830921345
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.5427392527377729
Number of packets orderedContinuous	0.590640086992017
Number of packets in orderedContinuous	0.590640086992017
Number of packets out orderedContinuous	0.992805755395683
Duration sortedContinuous	0.5645461820360992
Number control packets rate sortedContinuous	0.6700403827377984
Number data packets rate sortedContinuous	0.48499185999186
Number bytes transferred orderedContinuous	0.040021452983343
Number bytes transferred in orderedContinuous	0.040021452983343
Number bytes transferred out orderedContinuous	0.992805755395683
Number data bytes transferred orderedContinuous	0.0752407025042623
Number data bytes transferred in orderedContinuous	0.0752407025042623
Number data bytes transferred out orderedContinuous	0.992805755395683
Fragmented packets rate sortedContinuous	0.9880952380952381
Bad fragment rate sortedContinuous	0.9880952380952381
Max Src Window orderedContinuous	0.389136078401615
Max Dst Window orderedContinuous	0.9333333333333333
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	1
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.6697846446531464
Window exceeded rate sortedContinuous	1
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.520277847396855
Number reset connection orderedContinuous	0.813559322033898
Continued on next page	

Table L.32 – continued from previous page

Metric	Similarity
Number other errors orderedContinuous	0.813559322033898
Number disconnection errors orderedContinuous	0.813559322033898
Packet Destination IP nonKeyedSortedContinuous	0.785377011994492
Bytes Destination IP nonKeyedSortedContinuous	0.617622040427312
Connection Destination IP nonKeyedSortedContinuous	0.747266780703751
Packet Source IP nonKeyedSortedContinuous	0.689181181264378
Bytes Source IP nonKeyedSortedContinuous	0.572663849328549
Connection Source IP nonKeyedSortedContinuous	0.804491743496583

Table L.33: Similarity values for individual metrics of basecase 6,
test pair 2

Metric	Similarity
Packets in count	0.994021138942957
Packets out count	0.970272835522115
Connections in count	0.549208329620058
Connections out count	0.535663215695247
Bytes in count	0.969415763749213
Bytes out count	0.664186384337548
SYN-ONLY rate ratio	0.622002743484226
SYN-ACK rate ratio	0
Idle connection rate ratio	0.428549826100106
Half-open connection rate ratio	0.736952785817656
Packet Service discrete	0.0405511047329637
Bytes Service discrete	0.0371747429393052
Connection Service discrete	0.109326196197605
Packet Source port discrete	0.0384256398884685
Bytes Source port discrete	0.0350536894094428
Connection Source port discrete	0.0251563563976107
Continued on next page	

Table L.33 – continued from previous page

Metric	Similarity
Connection Source port orderedContinuous	0.0758443366759564
Packet TTL discrete	0.219273981247165
Packet TTL orderedContinuous	0.287992763411557
InterPacket delta sortedContinuous	0.7789767730303756
Packet sec orderedContinuous	0.93577351531408
Packet min orderedContinuous	0.467162815093753
Packet GmHour orderedContinuous	0.909603956612645
Packet LocHour orderedContinuous	0.909603956612645
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.914226092786231
Bytes min orderedContinuous	0.354659208005208
Bytes GmHour orderedContinuous	0.860112155249702
Bytes LocHour orderedContinuous	0.860112155249701
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.500673089582524
Packets in last w secs orderedContinuous	0.503744164033525
Priv packets time rate sortedContinuous	0.6116169823537902
Unpriv packets time rate sortedContinuous	0.9388401640562059
Connections time rate sortedContinuous	0.7593693337472735
Priv connections connection time rate sortedContinuous	0.9280441071710167
Unpriv connections connection time rate sortedContinuous	0.6675400784662797
Priv packets priv connection time rate sortedContinuous	0.7676871129292695
Unpriv packets unpriv connection time rate sortedContinuous	0.4643298796256695
SYNs connection time rate sortedContinuous	0.9569937959707487
RSTs connection time rate sortedContinuous	0.6310365418517361
FINs connection time rate sortedContinuous	0.8975117574829519
PSH connection time rate sortedContinuous	0.7803657043835026
Establishment errors connection time rate sortedContinuous	0.4841232176830062
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9723815083084022
Ave duration over last w secs sortedContinuous	0.8765413753924006
Continued on next page	

Table L.33 – continued from previous page

Metric	Similarity
Priv packets packet rate sortedContinuous	0.5861090342883659
Unpriv packets packet rate sortedContinuous	0.9404199221282551
InterConnection delta sortedContinuous	0.2781535568842545
Connection sec orderedContinuous	0.771608099365966
Connection min orderedContinuous	0.747176173677259
Connection GmHour orderedContinuous	0.949141816910482
Connection LocHour orderedContinuous	0.949141816910482
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.1990680369250818
Connection Priv connections rate sortedContinuous	0.668750361372575
Connection Unpriv connections rate sortedContinuous	0.1788303236901946
Connection Priv packet rate sortedContinuous	0.08953141478230035
Connection Unpriv packet rate sortedContinuous	0.1821393909190142
Connection SYNs rate sortedContinuous	0.1789809005311963
Connection RSTs rate sortedContinuous	0.6230982133188887
Connection FINs rate sortedContinuous	0.4623958132565157
Connection PSH rate sortedContinuous	0.1541723726202567
Connection Establishment errors rate sortedContinuous	0.7490632000226314
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9988177763447794
Ave duration over last m connections sortedContinuous	0.1961662688491047
Number of packets orderedContinuous	0.13714731927413
Number of packets in orderedContinuous	0.135080995617633
Number of packets out orderedContinuous	0.132204498950205
Duration sortedContinuous	0.9269542448047053
Number control packets rate sortedContinuous	0.9439023428138951
Number data packets rate sortedContinuous	0.9614528899502573
Number bytes transferred orderedContinuous	0.0650857923428926
Number bytes transferred in orderedContinuous	0.0780068383112597
Number bytes transferred out orderedContinuous	0.098363252315417
Number data bytes transferred orderedContinuous	0.044464846599231
Continued on next page	

Table L.33 – continued from previous page

Metric	Similarity
Number data bytes transferred in orderedContinuous	0.0478317491503221
Number data bytes transferred out orderedContinuous	0.0774267376041044
Fragmented packets rate sortedContinuous	0.9982215994826471
Bad fragment rate sortedContinuous	0.9982215994826471
Max Src Window orderedContinuous	0.115807988686019
Max Dst Window orderedContinuous	0.0910426476249499
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9895865090583343
Wrong resend rate sortedContinuous	0.9857119009908236
Duplicate ACK rate sortedContinuous	0.9992301770592764
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9852718679236918
Hole rate sortedContinuous	0.9977897363079916
Number connection errors orderedContinuous	0.30342003926568
Number reset connection orderedContinuous	0.517046102531252
Number other errors orderedContinuous	0.508631816390013
Number disconnection errors orderedContinuous	0.508643960371205
Packet Destination IP nonKeyedSortedContinuous	0.792899822160085
Bytes Destination IP nonKeyedSortedContinuous	0.734351999082742
Connection Destination IP nonKeyedSortedContinuous	0.775717279917436
Packet Source IP nonKeyedSortedContinuous	0.858973120695389
Bytes Source IP nonKeyedSortedContinuous	0.820525212754047
Connection Source IP nonKeyedSortedContinuous	0.890375097912848

Table L.34: Similarity values for individual metrics of basecase 6,
test pair 3

Metric	Similarity
Packets in count	0.284705924505993
Packets out count	0.378525095459355
Connections in count	0.766144200626959
Connections out count	0.781911013858497
Bytes in count	0.0947202976400754
Bytes out count	0.784176781422152
SYN-ONLY rate ratio	0.675869685256764
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.361341323473161
Packet Service discrete	0.0111771093441557
Bytes Service discrete	0.00903893111122018
Connection Service discrete	0.150136708181632
Packet Source port discrete	0.0143840747491227
Bytes Source port discrete	0.0122531187870821
Connection Source port discrete	0.0227951562133166
Connection Source port orderedContinuous	0.0407262228489961
Packet TTL discrete	0.14315774167543
Packet TTL orderedContinuous	0.230325489031497
InterPacket delta sortedContinuous	0.6059987242576854
Packet sec orderedContinuous	0.715129547642764
Packet min orderedContinuous	0.728719258836152
Packet GmHour orderedContinuous	0.929723599560063
Packet LocHour orderedContinuous	0.929723599560063
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.551038968590785
Bytes min orderedContinuous	0.526965328007336
Bytes GmHour orderedContinuous	0.783751845396007
Bytes LocHour orderedContinuous	0.783751845396007
Continued on next page	

Table L.34 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.286390587490147
Packets in last w secs orderedContinuous	0.1076498237415
Priv packets time rate sortedContinuous	0.7619575953080082
Unpriv packets time rate sortedContinuous	0.8661315199596919
Connections time rate sortedContinuous	0.3516219521786333
Priv connections connection time rate sortedContinuous	0.9567839684528839
Unpriv connections connection time rate sortedContinuous	0.8185264253351774
Priv packets priv connection time rate sortedContinuous	0.2205259019423339
Unpriv packets unpriv connection time rate sortedContinuous	0.5469470050503716
SYNs connection time rate sortedContinuous	0.3149436975181001
RSTs connection time rate sortedContinuous	0.7842087034012575
FINs connection time rate sortedContinuous	0.2820449468853908
PSH connection time rate sortedContinuous	0.4947268559219661
Establishment errors connection time rate sortedContinuous	0.9570669449823199
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.3092005296908377
Priv packets packet rate sortedContinuous	0.6839858225205876
Unpriv packets packet rate sortedContinuous	0.8246582568834331
InterConnection delta sortedContinuous	0.6750229525078394
Connection sec orderedContinuous	0.857542673525821
Connection min orderedContinuous	0.848424622294684
Connection GmHour orderedContinuous	0.991206203393361
Connection LocHour orderedContinuous	0.991206203393361
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.1596299444288187
Connection Priv connections rate sortedContinuous	0.8475839339838849
Connection Unpriv connections rate sortedContinuous	0.4528492018637339
Connection Priv packet rate sortedContinuous	0.07361939863455021
Connection Unpriv packet rate sortedContinuous	0.2979383989234417
Continued on next page	

Table L.34 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.6463242520101742
Connection RSTs rate sortedContinuous	0.4810204126496285
Connection FINs rate sortedContinuous	0.5414090696120728
Connection PSH rate sortedContinuous	0.2894471517273179
Connection Establishment errors rate sortedContinuous	0.9176118026740434
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.3902329949545498
Number of packets orderedContinuous	0.182317323274039
Number of packets in orderedContinuous	0.188177912422359
Number of packets out orderedContinuous	0.203713485190802
Duration sortedContinuous	0.899475400202677
Number control packets rate sortedContinuous	0.8284516491114374
Number data packets rate sortedContinuous	0.7956711924213542
Number bytes transferred orderedContinuous	0.0967269944687536
Number bytes transferred in orderedContinuous	0.0565367435622839
Number bytes transferred out orderedContinuous	0.073858250702983
Number data bytes transferred orderedContinuous	0.0134421955534809
Number data bytes transferred in orderedContinuous	0.0149809360156999
Number data bytes transferred out orderedContinuous	0.0318513540473481
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0052666227781435
Max Dst Window orderedContinuous	0.0376848857896035
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9656862745098039
Wrong resend rate sortedContinuous	0.9803921568627451
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9901960784313726
Continued on next page	

Table L.34 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9950980392156863
Number connection errors orderedContinuous	0.671882572797322
Number reset connection orderedContinuous	0.863095234991701
Number other errors orderedContinuous	0.941923288161138
Number disconnection errors orderedContinuous	0.941923288161138
Packet Destination IP nonKeyedSortedContinuous	0.586705906580754
Bytes Destination IP nonKeyedSortedContinuous	0.613186418736135
Connection Destination IP nonKeyedSortedContinuous	0.7836370735883
Packet Source IP nonKeyedSortedContinuous	0.695230696487723
Bytes Source IP nonKeyedSortedContinuous	0.757365506082313
Connection Source IP nonKeyedSortedContinuous	0.680166164430298

Table L.35: Similarity values for individual metrics of basecase 7,
train pair 1

Metric	Similarity
Packets in count	0.382218458933108
Packets out count	0.231190678293078
Connections in count	0.552147239263804
Connections out count	0.816326530612245
Bytes in count	0.819906339278368
Bytes out count	0.0304682971760916
SYN-ONLY rate ratio	0.923076923076922
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.84507042253521
Packet Service discrete	0.00382165492946904
Bytes Service discrete	0.00476283598716409
Connection Service discrete	0.00900130052257784
Continued on next page	

Table L.35 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.00252714758045197
Bytes Source port discrete	0.00183268583199054
Connection Source port discrete	0.00257029083974804
Connection Source port orderedContinuous	0.00426795207340877
Packet TTL discrete	0.114477098956437
Packet TTL orderedContinuous	0.175293110607682
InterPacket delta sortedContinuous	0.4305869952305174
Packet sec orderedContinuous	0.514030076900231
Packet min orderedContinuous	0.262940539891638
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.373794292496319
Bytes min orderedContinuous	0.21240859599239
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.145396261120238
Packets in last w secs orderedContinuous	0.357841121683753
Priv packets time rate sortedContinuous	0.8822653492884632
Unpriv packets time rate sortedContinuous	0.9565045481895901
Connections time rate sortedContinuous	0.05963198981375069
Priv connections connection time rate sortedContinuous	0.7483896601692964
Unpriv connections connection time rate sortedContinuous	0.6021304926764314
Priv packets priv connection time rate sortedContinuous	0.07160207382378531
Unpriv packets unpriv connection time rate sortedContinuous	0.6021304926764314
SYNs connection time rate sortedContinuous	0.1152470983365108
RSTs connection time rate sortedContinuous	0.7545344128431371
FINs connection time rate sortedContinuous	0.3828725950660704
PSH connection time rate sortedContinuous	0.01733407923030728
Establishment errors connection time rate sortedContinuous	0.7117768900724959
Continued on next page	

Table L.35 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.0696919936304426
Priv packets packet rate sortedContinuous	0.9212841539420981
Unpriv packets packet rate sortedContinuous	0.9667020283640172
InterConnection delta sortedContinuous	0.2022045152131581
Connection sec orderedContinuous	0.431589572049926
Connection min orderedContinuous	0.288361292722167
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.3290874447326106
Connection Priv connections rate sortedContinuous	0.6881251696172431
Connection Unpriv connections rate sortedContinuous	0
Connection Priv packet rate sortedContinuous	0.2774832242927514
Connection Unpriv packet rate sortedContinuous	0
Connection SYNs rate sortedContinuous	0.5025494371275259
Connection RSTs rate sortedContinuous	0.1181185959904061
Connection FINs rate sortedContinuous	0.4636944489217578
Connection PSH rate sortedContinuous	0.2195279071563061
Connection Establishment errors rate sortedContinuous	0.1140486647920985
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.1021970604125773
Number of packets orderedContinuous	0.163975582699734
Number of packets in orderedContinuous	0.202946459233977
Number of packets out orderedContinuous	0.179717016309096
Duration sortedContinuous	0.8405876104675076
Number control packets rate sortedContinuous	0.8767155747780278
Number data packets rate sortedContinuous	0.9120516572773945
Number bytes transferred orderedContinuous	0.0888745512088029
Continued on next page	

Table L.35 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.081092720208634
Number bytes transferred out orderedContinuous	0.0804182832383585
Number data bytes transferred orderedContinuous	0.0121513563102562
Number data bytes transferred in orderedContinuous	0.0135015070113958
Number data bytes transferred out orderedContinuous	0.0167387766213156
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.121023903632599
Max Dst Window orderedContinuous	0.0827894194434902
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9873294346978558
Wrong resend rate sortedContinuous	0.9444444444444444
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9444444444444444
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.3978196159871
Number reset connection orderedContinuous	0.569794853701564
Number other errors orderedContinuous	0.585690515806988
Number disconnection errors orderedContinuous	0.585690515806988
Packet Destination IP nonKeyedSortedContinuous	0.625703768237368
Bytes Destination IP nonKeyedSortedContinuous	0.631492867908503
Connection Destination IP nonKeyedSortedContinuous	0.488036809815951
Packet Source IP nonKeyedSortedContinuous	0.612480708541556
Bytes Source IP nonKeyedSortedContinuous	0.606523281010893
Connection Source IP nonKeyedSortedContinuous	0.741045066045066

Table L.36: Similarity values for individual metrics of basecase 7,
train pair 2

Metric	Similarity
Packets in count	0.347306417175413
Packets out count	0.381046370861624
Connections in count	0.531634381674891
Connections out count	0.84647619047619
Bytes in count	0.249102577412536
Bytes out count	0.68038329388676
SYN-ONLY rate ratio	0.749748409759802
SYN-ACK rate ratio	0.371629109715551
Idle connection rate ratio	0.184181618454779
Half-open connection rate ratio	0.67278501718253
Packet Service discrete	0.0208922085444719
Bytes Service discrete	0.0163051429991655
Connection Service discrete	0.00551215364656296
Packet Source port discrete	0.0254197836038913
Bytes Source port discrete	0.0212521190960625
Connection Source port discrete	0.0437686764168344
Connection Source port orderedContinuous	0.0811268869098638
Packet TTL discrete	0.140613588352523
Packet TTL orderedContinuous	0.23533212558524
InterPacket delta sortedContinuous	0.5302145450748991
Packet sec orderedContinuous	0.912263516934939
Packet min orderedContinuous	0.718633085994397
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.867188360291619
Bytes min orderedContinuous	0.582994137976967
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Continued on next page	

Table L.36 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.46192154675836
Packets in last w secs orderedContinuous	0.089662441396758
Priv packets time rate sortedContinuous	0.6632507736495353
Unpriv packets time rate sortedContinuous	0.7146773940093186
Connections time rate sortedContinuous	0.3356385813761609
Priv connections connection time rate sortedContinuous	0.9827864012374288
Unpriv connections connection time rate sortedContinuous	0.6466288307072445
Priv packets priv connection time rate sortedContinuous	0.6060716764116329
Unpriv packets unpriv connection time rate sortedContinuous	0.3947612983911302
SYNs connection time rate sortedContinuous	0.7804525180269751
RSTs connection time rate sortedContinuous	0.6497593720259168
FINs connection time rate sortedContinuous	0.8421419830774767
PSH connection time rate sortedContinuous	0.2476975227895033
Establishment errors connection time rate sortedContinuous	0.7076690991169835
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9884555178711192
Ave duration over last w secs sortedContinuous	0.1427150050562767
Priv packets packet rate sortedContinuous	0.6342651134131245
Unpriv packets packet rate sortedContinuous	0.708894041570107
InterConnection delta sortedContinuous	0.2202301139525702
Connection sec orderedContinuous	0.957336665508498
Connection min orderedContinuous	0.568179166154748
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.04345042287976284
Connection Priv connections rate sortedContinuous	0.7780818349023472
Connection Unpriv connections rate sortedContinuous	0.7092015571875492
Connection Priv packet rate sortedContinuous	0.5745995143717334
Connection Unpriv packet rate sortedContinuous	0.0396669080664577
Continued on next page	

Table L.36 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.9244913148017719
Connection RSTs rate sortedContinuous	0.7073993784561744
Connection FINs rate sortedContinuous	0.6895572948714256
Connection PSH rate sortedContinuous	0.03412724295858809
Connection Establishment errors rate sortedContinuous	0.7413694830472945
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.998035792230467
Ave duration over last m connections sortedContinuous	0.3073673030446833
Number of packets orderedContinuous	0.382775839509926
Number of packets in orderedContinuous	0.327801416637871
Number of packets out orderedContinuous	0.345666891125177
Duration sortedContinuous	0.8958016986563496
Number control packets rate sortedContinuous	0.954448730554069
Number data packets rate sortedContinuous	0.9121710031215105
Number bytes transferred orderedContinuous	0.0772096486236579
Number bytes transferred in orderedContinuous	0.0856890555138429
Number bytes transferred out orderedContinuous	0.173739458912223
Number data bytes transferred orderedContinuous	0.0567516529543781
Number data bytes transferred in orderedContinuous	0.0553626836461889
Number data bytes transferred out orderedContinuous	0.18116915552191
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.282783842838157
Max Dst Window orderedContinuous	0.0768175739350512
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.97200381941495
Wrong resend rate sortedContinuous	0.9976027188558094
Duplicate ACK rate sortedContinuous	0.9993373094764745
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9176206277933265
Window exceeded rate sortedContinuous	0.9971437869379947
Continued on next page	

Table L.36 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9991507413637681
Number connection errors orderedContinuous	0.413721195955378
Number reset connection orderedContinuous	0.917572216204285
Number other errors orderedContinuous	0.932500480171024
Number disconnection errors orderedContinuous	0.812497338742164
Packet Destination IP nonKeyedSortedContinuous	0.799979430976489
Bytes Destination IP nonKeyedSortedContinuous	0.768238370037387
Connection Destination IP nonKeyedSortedContinuous	0.88967475549729
Packet Source IP nonKeyedSortedContinuous	0.776907076468389
Bytes Source IP nonKeyedSortedContinuous	0.698876818983215
Connection Source IP nonKeyedSortedContinuous	0.655494779996587

Table L.37: Similarity values for individual metrics of basecase 7,
train pair 3

Metric	Similarity
Packets in count	0.000135873257425478
Packets out count	0.000122846349927874
Connections in count	0.000292611558156519
Connections out count	0.000146412884333813
Bytes in count	1.35033255137529e-05
Bytes out count	0.000116397719182704
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	3.44146592006924e-06
Bytes Service discrete	3.92994801778812e-07
Connection Service discrete	1.44558032543231e-05
Continued on next page	

Table L.37 – continued from previous page

Metric	Similarity
Packet Source port discrete	5.05671369092242e-06
Bytes Source port discrete	7.74963413576897e-07
Connection Source port discrete	8.69929605893553e-06
Connection Source port orderedContinuous	6.88199378877521e-05
Packet TTL discrete	0.00409831295065146
Packet TTL orderedContinuous	0.0451523458433014
InterPacket delta sortedContinuous	0.5779927610276609
Packet sec orderedContinuous	0.0152431642087191
Packet min orderedContinuous	0.00809046881740436
Packet GmHour orderedContinuous	0.4166666666666667
Packet LocHour orderedContinuous	0.238095238095238
Packet weekday orderedContinuous	0.333333333333333
Bytes sec orderedContinuous	0.0148655479651693
Bytes min orderedContinuous	0.00824806347245917
Bytes GmHour orderedContinuous	0.4166666666666667
Bytes LocHour orderedContinuous	0.238095238095238
Bytes weekday orderedContinuous	0.333333333333333
Packet size orderedContinuous	0.0053921046894788
Packets in last w secs orderedContinuous	5.66021006071477e-07
Priv packets time rate sortedContinuous	0.3657113376928271
Unpriv packets time rate sortedContinuous	0.6964547522189696
Connections time rate sortedContinuous	0.1553536978117283
Priv connections connection time rate sortedContinuous	0.7170417165586248
Unpriv connections connection time rate sortedContinuous	0.1111111111111111
Priv packets priv connection time rate sortedContinuous	0.2975591344847395
Unpriv packets unpriv connection time rate sortedContinuous	0.1111111111111111
SYNs connection time rate sortedContinuous	0.7754232412666346
RSTs connection time rate sortedContinuous	0.4788213627992633
FINs connection time rate sortedContinuous	0.2222222222222222
PSH connection time rate sortedContinuous	0.2429909052581393
Establishment errors connection time rate sortedContinuous	0.5555555555555556

Continued on next page

Table L.37 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.8888888888888888
Ave duration over last w secs sortedContinuous	0.1112266606229539
Priv packets packet rate sortedContinuous	0.3400940541322151
Unpriv packets packet rate sortedContinuous	0.698171119558103
InterConnection delta sortedContinuous	0.6656067732143898
Connection sec orderedContinuous	0.0081239852926626
Connection min orderedContinuous	0.00197565228579508
Connection GmHour orderedContinuous	0.4166666666666667
Connection LocHour orderedContinuous	0.238095238095238
Connection weekday orderedContinuous	0.3333333333333333
Connection packet rate sortedContinuous	0.4192669500448183
Connection Priv connections rate sortedContinuous	0.358974358974359
Connection Unpriv connections rate sortedContinuous	0.3333333333333333
Connection Priv packet rate sortedContinuous	0.3396085462423427
Connection Unpriv packet rate sortedContinuous	0.3333333333333333
Connection SYNs rate sortedContinuous	0.4328358208955223
Connection RSTs rate sortedContinuous	0.6666666666666666
Connection FINs rate sortedContinuous	0.6666666666666666
Connection PSH rate sortedContinuous	0.3333333333333333
Connection Establishment errors rate sortedContinuous	0.6666666666666666
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.6666666666666666
Ave duration over last m connections sortedContinuous	0.3333333333333333
Number of packets orderedContinuous	0.0107919296402239
Number of packets in orderedContinuous	0.0200655247540922
Number of packets out orderedContinuous	0.0188319597694767
Duration sortedContinuous	0.6667371820887112
Number control packets rate sortedContinuous	0.9090909090909089
Number data packets rate sortedContinuous	0.6666666666666666
Number bytes transferred orderedContinuous	9.78133309234308e-05

Continued on next page

Table L.37 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.000798007164871863
Number bytes transferred out orderedContinuous	0.000930440502685393
Number data bytes transferred orderedContinuous	0.000798368450063495
Number data bytes transferred in orderedContinuous	0.000907570908896712
Number data bytes transferred out orderedContinuous	0.00131058994786901
Fragmented packets rate sortedContinuous	0.666666666666666666
Bad fragment rate sortedContinuous	0.666666666666666666
Max Src Window orderedContinuous	0
Max Dst Window orderedContinuous	0.00154918667699458
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	1
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	1
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.0647481907943542
Number reset connection orderedContinuous	0.120304969373488
Number other errors orderedContinuous	0.176741393114492
Number disconnection errors orderedContinuous	0.152220993474833
Packet Destination IP nonKeyedSortedContinuous	0.200150756799132
Bytes Destination IP nonKeyedSortedContinuous	0.103792209981538
Connection Destination IP nonKeyedSortedContinuous	0.500210150257434
Packet Source IP nonKeyedSortedContinuous	0.20015675455372
Bytes Source IP nonKeyedSortedContinuous	0.112168458915979
Connection Source IP nonKeyedSortedContinuous	0.50021215657155

Table L.38: Similarity values for individual metrics of basecase 7,
test pair 1

Metric	Similarity
Packets in count	0.769938650306748
Packets out count	0.891891891891892
Connections in count	0.80952380952381
Connections out count	0.567901234567901
Bytes in count	0.474926945360604
Bytes out count	0.782809541840212
SYN-ONLY rate ratio	0.348122866894198
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.392307692307692
Packet Service discrete	0.00244210869503954
Bytes Service discrete	0.00315619559094604
Connection Service discrete	0.00240164787496567
Packet Source port discrete	0.000856206662161351
Bytes Source port discrete	0.000619412995125756
Connection Source port discrete	0.00174278120373902
Connection Source port orderedContinuous	0.00426298921005401
Packet TTL discrete	0.0763035859751459
Packet TTL orderedContinuous	0.170706981020997
InterPacket delta sortedContinuous	0.620755419645936
Packet sec orderedContinuous	0.465859266302426
Packet min orderedContinuous	0.401364271165837
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.379378504707925
Bytes min orderedContinuous	0.273738560214409
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Continued on next page	

Table L.38 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.182012738147161
Packets in last w secs orderedContinuous	0.358868526100616
Priv packets time rate sortedContinuous	0.8212284982198739
Unpriv packets time rate sortedContinuous	0.9116703338328202
Connections time rate sortedContinuous	0.7385151980869688
Priv connections connection time rate sortedContinuous	0.6924070223063046
Unpriv connections connection time rate sortedContinuous	0.480054090601758
Priv packets priv connection time rate sortedContinuous	0.8068013292030209
Unpriv packets unpriv connection time rate sortedContinuous	0.480054090601758
SYNs connection time rate sortedContinuous	0.7903761217979985
RSTs connection time rate sortedContinuous	0.8281153928154794
FINs connection time rate sortedContinuous	0.7412651370998307
PSH connection time rate sortedContinuous	0.7304627088554412
Establishment errors connection time rate sortedContinuous	0.9783022927039154
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.3622184153287358
Priv packets packet rate sortedContinuous	0.9040990031516288
Unpriv packets packet rate sortedContinuous	0.948903901520833
InterConnection delta sortedContinuous	0.2083136822555637
Connection sec orderedContinuous	0.522326677092428
Connection min orderedContinuous	0.477197021720413
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.6100913209376129
Connection Priv connections rate sortedContinuous	0.7584549831213074
Connection Unpriv connections rate sortedContinuous	0.007233721272681821
Connection Priv packet rate sortedContinuous	0.4627101218578093
Connection Unpriv packet rate sortedContinuous	0.06148148148148148
Continued on next page	

Table L.38 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.441737111806378
Connection RSTs rate sortedContinuous	0.223487748541784
Connection FINs rate sortedContinuous	0.3471749317320843
Connection PSH rate sortedContinuous	0.4350480925498076
Connection Establishment errors rate sortedContinuous	0.7557037037037037
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.5527947405871761
Number of packets orderedContinuous	0.0923868942219826
Number of packets in orderedContinuous	0.229408361991302
Number of packets out orderedContinuous	0.203986381619818
Duration sortedContinuous	0.9240123928277101
Number control packets rate sortedContinuous	0.9130052028457195
Number data packets rate sortedContinuous	0.9483424245253526
Number bytes transferred orderedContinuous	0.09975001468988
Number bytes transferred in orderedContinuous	0.0936277691107171
Number bytes transferred out orderedContinuous	0.0731262095959067
Number data bytes transferred orderedContinuous	0.0138131948281357
Number data bytes transferred in orderedContinuous	0.0289395080110995
Number data bytes transferred out orderedContinuous	0.0293921243315472
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0234113712374582
Max Dst Window orderedContinuous	0.0399305555555556
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.8070961707466634
Wrong resend rate sortedContinuous	0.9878787878787878
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9249374478732277
Continued on next page	

Table L.38 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9545454545454546
Number connection errors orderedContinuous	0.378748617281706
Number reset connection orderedContinuous	0.374179168610828
Number other errors orderedContinuous	0.378722671716302
Number disconnection errors orderedContinuous	0.378722671716302
Packet Destination IP nonKeyedSortedContinuous	0.605229473098017
Bytes Destination IP nonKeyedSortedContinuous	0.570039233811205
Connection Destination IP nonKeyedSortedContinuous	0.707888908317305
Packet Source IP nonKeyedSortedContinuous	0.549229714074458
Bytes Source IP nonKeyedSortedContinuous	0.698436258561718
Connection Source IP nonKeyedSortedContinuous	0.743931744676914

Table L.39: Similarity values for individual metrics of basecase 7,
test pair 2

Metric	Similarity
Packets in count	0.663213448006255
Packets out count	0.659261654335546
Connections in count	0.255221386800334
Connections out count	0.671491757137113
Bytes in count	0.571259225200471
Bytes out count	0.885264084965119
SYN-ONLY rate ratio	0.606343742316203
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0.928868388938301
Packet Service discrete	0.0098880990864844
Bytes Service discrete	0.00832272026170131
Connection Service discrete	0.0347988948350439
Continued on next page	

Table L.39 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.0100146973423261
Bytes Source port discrete	0.00803037186256186
Connection Source port discrete	0.00898162507185469
Connection Source port orderedContinuous	0.021699318886847
Packet TTL discrete	0.186072171777241
Packet TTL orderedContinuous	0.237220838045059
InterPacket delta sortedContinuous	0.6517650531700841
Packet sec orderedContinuous	0.92448957535698
Packet min orderedContinuous	0.627302088086815
Packet GmHour orderedContinuous	0.951365751603022
Packet LocHour orderedContinuous	0.951365751603022
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.899999032249434
Bytes min orderedContinuous	0.57722328307024
Bytes GmHour orderedContinuous	0.89435558656192
Bytes LocHour orderedContinuous	0.89435558656192
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.373873795639881
Packets in last w secs orderedContinuous	0.156596250450224
Priv packets time rate sortedContinuous	0.5578814670869352
Unpriv packets time rate sortedContinuous	0.846101716468374
Connections time rate sortedContinuous	0.4009509899514277
Priv connections connection time rate sortedContinuous	0.7867033085877039
Unpriv connections connection time rate sortedContinuous	0.3097189692166277
Priv packets priv connection time rate sortedContinuous	0.2312831574371505
Unpriv packets unpriv connection time rate sortedContinuous	0.1171151631778402
SYNs connection time rate sortedContinuous	0.4761461338801363
RSTs connection time rate sortedContinuous	0.7059186580911191
FINs connection time rate sortedContinuous	0.5238151993596262
PSH connection time rate sortedContinuous	0.3662647074099531
Establishment errors connection time rate sortedContinuous	0.387629912077497
Continued on next page	

Table L.39 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.3660398661349007
Priv packets packet rate sortedContinuous	0.5155075745258474
Unpriv packets packet rate sortedContinuous	0.8552948358540325
InterConnection delta sortedContinuous	0.5293550179964871
Connection sec orderedContinuous	0.855231711410516
Connection min orderedContinuous	0.657903981047156
Connection GmHour orderedContinuous	0.879354552655091
Connection LocHour orderedContinuous	0.879354552655091
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.2316933775875603
Connection Priv connections rate sortedContinuous	0.5735159535971653
Connection Unpriv connections rate sortedContinuous	0.4073750862849243
Connection Priv packet rate sortedContinuous	0.04567598926132995
Connection Unpriv packet rate sortedContinuous	0.1611967807012974
Connection SYNs rate sortedContinuous	0.7917758408780439
Connection RSTs rate sortedContinuous	0.6587708354055066
Connection FINs rate sortedContinuous	0.7446467882989102
Connection PSH rate sortedContinuous	0.2663805392928644
Connection Establishment errors rate sortedContinuous	0.6978088033067286
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.8587736670486558
Number of packets orderedContinuous	0.233191672116872
Number of packets in orderedContinuous	0.27646455047835
Number of packets out orderedContinuous	0.292170834957286
Duration sortedContinuous	0.8567588500902479
Number control packets rate sortedContinuous	0.7451590989071605
Number data packets rate sortedContinuous	0.7765899203365485
Number bytes transferred orderedContinuous	0.0587120006191651
Continued on next page	

Table L.39 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.0558272497756005
Number bytes transferred out orderedContinuous	0.094981410949823
Number data bytes transferred orderedContinuous	0.0300239011742375
Number data bytes transferred in orderedContinuous	0.0245429903468198
Number data bytes transferred out orderedContinuous	0.0917828416940014
Fragmented packets rate sortedContinuous	0.5342756183745583
Bad fragment rate sortedContinuous	0.5342756183745583
Max Src Window orderedContinuous	0.0680583661731814
Max Dst Window orderedContinuous	0.0931611935992604
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.8808789680581882
Wrong resend rate sortedContinuous	0.8983886173108063
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9908256880733946
Hole rate sortedContinuous	0.9954128440366973
Number connection errors orderedContinuous	0.255763491940136
Number reset connection orderedContinuous	0.633190185417326
Number other errors orderedContinuous	0.637315900887279
Number disconnection errors orderedContinuous	0.637315900887279
Packet Destination IP nonKeyedSortedContinuous	0.764727512291156
Bytes Destination IP nonKeyedSortedContinuous	0.780041462546352
Connection Destination IP nonKeyedSortedContinuous	0.806719001762723
Packet Source IP nonKeyedSortedContinuous	0.737616085157659
Bytes Source IP nonKeyedSortedContinuous	0.734464780146914
Connection Source IP nonKeyedSortedContinuous	0.826166771926641

Table L.40: Similarity values for individual metrics of basecase 7,
test pair 3

Metric	Similarity
Packets in count	0
Packets out count	0.00335494616620069
Connections in count	0
Connections out count	0.0815217391304348
Bytes in count	0
Bytes out count	0.00137903858324606
SYN-ONLY rate ratio	1
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	1
Packet Service discrete	0.0563673258843082
Bytes Service discrete	0.0411578853855403
Connection Service discrete	0.0408472005121503
Packet Source port discrete	0.0216401962565641
Bytes Source port discrete	0.0168437263304481
Connection Source port discrete	0.0103624581246134
Connection Source port orderedContinuous	0.0241734416966101
Packet TTL discrete	0.135827809213321
Packet TTL orderedContinuous	0.176896868077066
InterPacket delta sortedContinuous	0.5938686103760722
Packet sec orderedContinuous	0.969894415083994
Packet min orderedContinuous	0.875367367323222
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.930600398110858
Bytes min orderedContinuous	0.818291027187053
Bytes GmHour orderedContinuous	0.999655121453983
Bytes LocHour orderedContinuous	0.999655121453983
Continued on next page	

Table L.40 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	0.999655121453983
Packet size orderedContinuous	0.419248360038747
Packets in last w secs orderedContinuous	0.26279968661853
Priv packets time rate sortedContinuous	0.6754274386449702
Unpriv packets time rate sortedContinuous	0.9912401368767298
Connections time rate sortedContinuous	0.8544383918285248
Priv connections connection time rate sortedContinuous	0.8428609206581387
Unpriv connections connection time rate sortedContinuous	0.5137364744336245
Priv packets priv connection time rate sortedContinuous	0.5418796848224806
Unpriv packets unpriv connection time rate sortedContinuous	0.4066243411220027
SYNs connection time rate sortedContinuous	0.6679925734792613
RSTs connection time rate sortedContinuous	0.7716143050368276
FINs connection time rate sortedContinuous	0.6518511644694419
PSH connection time rate sortedContinuous	0.9646829137292836
Establishment errors connection time rate sortedContinuous	0.6871633473274637
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.8297868665525775
Priv packets packet rate sortedContinuous	0.9725971649724681
Unpriv packets packet rate sortedContinuous	0.9965311896705551
InterConnection delta sortedContinuous	0.6040154573194037
Connection sec orderedContinuous	0.849202481937287
Connection min orderedContinuous	0.764067063036944
Connection GmHour orderedContinuous	0.979195561719834
Connection LocHour orderedContinuous	0.979195561719834
Connection weekday orderedContinuous	0.979195561719834
Connection packet rate sortedContinuous	0.8091828479713581
Connection Priv connections rate sortedContinuous	0.8664116156658261
Connection Unpriv connections rate sortedContinuous	0.2455508360280076
Connection Priv packet rate sortedContinuous	0.9304298323268175
Connection Unpriv packet rate sortedContinuous	0.3901323193494167
Continued on next page	

Table L.40 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.6908378729626633
Connection RSTs rate sortedContinuous	0.560910782422901
Connection FINs rate sortedContinuous	0.4655181613649191
Connection PSH rate sortedContinuous	0.6521974192406605
Connection Establishment errors rate sortedContinuous	0.4177713196247317
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.5296388975643698
Number of packets orderedContinuous	0.389948439586517
Number of packets in orderedContinuous	0.0803771581674288
Number of packets out orderedContinuous	0.394272366510593
Duration sortedContinuous	0.9084691928859422
Number control packets rate sortedContinuous	0.9295289289616203
Number data packets rate sortedContinuous	0.928763653855503
Number bytes transferred orderedContinuous	0.139904301099127
Number bytes transferred in orderedContinuous	0.0011191060629426
Number bytes transferred out orderedContinuous	0.0372907562358379
Number data bytes transferred orderedContinuous	0.0254529201041975
Number data bytes transferred in orderedContinuous	0.0314622866296841
Number data bytes transferred out orderedContinuous	0.0359613812822475
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.227305472882996
Max Dst Window orderedContinuous	0.8444444444444444
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	1
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.3383178760847337
Window exceeded rate sortedContinuous	1
Continued on next page	

Table L.40 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.372429039008819
Number reset connection orderedContinuous	0.813809483416744
Number other errors orderedContinuous	0.813809483416744
Number disconnection errors orderedContinuous	0.813809483416744
Packet Destination IP nonKeyedSortedContinuous	0.822406301272383
Bytes Destination IP nonKeyedSortedContinuous	0.670732261161947
Connection Destination IP nonKeyedSortedContinuous	0.690251133008208
Packet Source IP nonKeyedSortedContinuous	0.126287249514402
Bytes Source IP nonKeyedSortedContinuous	0.118388305386669
Connection Source IP nonKeyedSortedContinuous	0.366929899622207

Table L.41: Similarity values for individual metrics of basecase 8,
train pair 1

Metric	Similarity
Packets in count	0.0149488388876002
Packets out count	0.0177227834436979
Connections in count	0.0578665367754505
Connections out count	0.0718288334182374
Bytes in count	0.00368189861262747
Bytes out count	0.0234031009119599
SYN-ONLY rate ratio	0.821110186615953
SYN-ACK rate ratio	0
Idle connection rate ratio	0.0262686567164179
Half-open connection rate ratio	0.869163673678811
Packet Service discrete	0.000260617531730439
Bytes Service discrete	0.000243273084910534
Connection Service discrete	0.000218215763799686
Continued on next page	

Table L.41 – continued from previous page

Metric	Similarity
Packet Source port discrete	8.32577326412232e-05
Bytes Source port discrete	7.3420751558706e-05
Connection Source port discrete	0.000166967296303745
Connection Source port orderedContinuous	0.00121771234626171
Packet TTL discrete	0.0360552155492959
Packet TTL orderedContinuous	0.0656786583451291
InterPacket delta sortedContinuous	0.6122414477057327
Packet sec orderedContinuous	0.662824082548825
Packet min orderedContinuous	0.335276443502083
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.621273350990556
Bytes min orderedContinuous	0.289624028078725
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.064871162073796
Packets in last w secs orderedContinuous	0.000516935692318725
Priv packets time rate sortedContinuous	0.5906473109373714
Unpriv packets time rate sortedContinuous	0.7759684904181972
Connections time rate sortedContinuous	0.7069695987702123
Priv connections connection time rate sortedContinuous	0.6208846296878067
Unpriv connections connection time rate sortedContinuous	0.02416680620904837
Priv packets priv connection time rate sortedContinuous	0.7871231195994591
Unpriv packets unpriv connection time rate sortedContinuous	0.0240167208366572
SYNs connection time rate sortedContinuous	0.657705261722755
RSTs connection time rate sortedContinuous	0.6202532030095331
FINs connection time rate sortedContinuous	0.5726243696216681
PSH connection time rate sortedContinuous	0.7808456653234645
Establishment errors connection time rate sortedContinuous	0.2665846295358377
Continued on next page	

Table L.41 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.03773399535728751
Priv packets packet rate sortedContinuous	0.5944079886803236
Unpriv packets packet rate sortedContinuous	0.8206583570656584
InterConnection delta sortedContinuous	0.6665757991461602
Connection sec orderedContinuous	0.583238497500145
Connection min orderedContinuous	0.244488474877239
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.03243255006757962
Connection Priv connections rate sortedContinuous	0.848248933914939
Connection Unpriv connections rate sortedContinuous	0.07367207196445976
Connection Priv packet rate sortedContinuous	0.4286514158529137
Connection Unpriv packet rate sortedContinuous	0.03791132243727964
Connection SYNs rate sortedContinuous	0.4388528223330388
Connection RSTs rate sortedContinuous	0.6922536641882572
Connection FINs rate sortedContinuous	0.8057391649720262
Connection PSH rate sortedContinuous	0.02105556319507156
Connection Establishment errors rate sortedContinuous	0.4118152218139515
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.08833625290493142
Number of packets orderedContinuous	0.0658774806912642
Number of packets in orderedContinuous	0.0849315424326812
Number of packets out orderedContinuous	0.0853433139889708
Duration sortedContinuous	0.8276721432735602
Number control packets rate sortedContinuous	0.8372287505931986
Number data packets rate sortedContinuous	0.8405498893343959
Number bytes transferred orderedContinuous	0.0210575934792346
Continued on next page	

Table L.41 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.0201270751043236
Number bytes transferred out orderedContinuous	0.0165657341973107
Number data bytes transferred orderedContinuous	0.00170769592377773
Number data bytes transferred in orderedContinuous	0.00220995944527654
Number data bytes transferred out orderedContinuous	0.00367802737772031
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0307742699743024
Max Dst Window orderedContinuous	0.00328692333369995
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9597268016957137
Wrong resend rate sortedContinuous	0.9268547544409614
Duplicate ACK rate sortedContinuous	0.9545454545454546
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9545454545454546
Window exceeded rate sortedContinuous	0.9545454545454546
Hole rate sortedContinuous	0.9714587737843552
Number connection errors orderedContinuous	0.0933200909027718
Number reset connection orderedContinuous	0.31820667561096
Number other errors orderedContinuous	0.335508106486494
Number disconnection errors orderedContinuous	0.335508106486494
Packet Destination IP nonKeyedSortedContinuous	0.540105105511022
Bytes Destination IP nonKeyedSortedContinuous	0.49834140805827
Connection Destination IP nonKeyedSortedContinuous	0.541931465899026
Packet Source IP nonKeyedSortedContinuous	0.283008374186569
Bytes Source IP nonKeyedSortedContinuous	0.198785044969257
Connection Source IP nonKeyedSortedContinuous	0.134743613408359

Table L.42: Similarity values for individual metrics of basecase 8,
train pair 2

Metric	Similarity
Packets in count	0.00454337043057818
Packets out count	0.00520685947139055
Connections in count	0.0876369327073553
Connections out count	0.0236686390532544
Bytes in count	0.00100268125792413
Bytes out count	0.00343763814788867
SYN-ONLY rate ratio	0.0792951541850219
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.604799999999999
Packet Service discrete	0.000159006011117204
Bytes Service discrete	3.57265351473561e-05
Connection Service discrete	0.0115809651160343
Packet Source port discrete	0.00212857383316633
Bytes Source port discrete	0.000716054699224216
Connection Source port discrete	0.00140654540835325
Connection Source port orderedContinuous	0.00642287423824793
Packet TTL discrete	0.0405194273356542
Packet TTL orderedContinuous	0.0866794514359439
InterPacket delta sortedContinuous	0.4036697338086017
Packet sec orderedContinuous	0.54267357388664
Packet min orderedContinuous	0.270617261899006
Packet GmHour orderedContinuous	0.749275020754935
Packet LocHour orderedContinuous	0.749275020754935
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.417917585402032
Bytes min orderedContinuous	0.244106610837462
Bytes GmHour orderedContinuous	0.748029396665034
Bytes LocHour orderedContinuous	0.748029396665034
Continued on next page	

Table L.42 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.0357926856757816
Packets in last w secs orderedContinuous	3.98436008125136e-05
Priv packets time rate sortedContinuous	0.8415228834190325
Unpriv packets time rate sortedContinuous	0.8634417499188708
Connections time rate sortedContinuous	0.1317775323877322
Priv connections connection time rate sortedContinuous	0.7739017322710466
Unpriv connections connection time rate sortedContinuous	0.478101555713496
Priv packets priv connection time rate sortedContinuous	0.1173149312486078
Unpriv packets unpriv connection time rate sortedContinuous	0.4454460435738153
SYNs connection time rate sortedContinuous	0.2833705306260382
RSTs connection time rate sortedContinuous	0.7512437810945274
FINs connection time rate sortedContinuous	0.7001202481162673
PSH connection time rate sortedContinuous	0.2328892678308562
Establishment errors connection time rate sortedContinuous	0.9439823378158021
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.2260761710653741
Priv packets packet rate sortedContinuous	0.9384809978365985
Unpriv packets packet rate sortedContinuous	0.9638392411563989
InterConnection delta sortedContinuous	0.3877191885176106
Connection sec orderedContinuous	0.496504994398878
Connection min orderedContinuous	0.365803631819613
Connection GmHour orderedContinuous	0.75960768506232
Connection LocHour orderedContinuous	0.75960768506232
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.02252391003779587
Connection Priv connections rate sortedContinuous	0.6994204597576584
Connection Unpriv connections rate sortedContinuous	0.7727142433108401
Connection Priv packet rate sortedContinuous	0.01410460978133063
Connection Unpriv packet rate sortedContinuous	0.8248768851954446

Continued on next page

Table L.42 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.4488336530822402
Connection RSTs rate sortedContinuous	0.6258297660902945
Connection FINs rate sortedContinuous	0.5592737809438004
Connection PSH rate sortedContinuous	0.02988229882069756
Connection Establishment errors rate sortedContinuous	0.2229130821956963
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.4730823010500614
Number of packets orderedContinuous	0.0563735755390803
Number of packets in orderedContinuous	0.109493133251473
Number of packets out orderedContinuous	0.08811011533917
Duration sortedContinuous	0.7576143671347429
Number control packets rate sortedContinuous	0.9349877664053859
Number data packets rate sortedContinuous	0.953407400439706
Number bytes transferred orderedContinuous	0.0149119121959978
Number bytes transferred in orderedContinuous	0.0110059712061659
Number bytes transferred out orderedContinuous	0.0145261562683078
Number data bytes transferred orderedContinuous	0.00492580807579526
Number data bytes transferred in orderedContinuous	0.00604723323393668
Number data bytes transferred out orderedContinuous	0.00629057535053058
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.00203252032520325
Max Dst Window orderedContinuous	0.0310880829015544
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9333333333333333
Wrong resend rate sortedContinuous	0.9333333333333333
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9333333333333333
Window exceeded rate sortedContinuous	0.9333333333333333
Continued on next page	

Table L.42 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9333333333333333
Number connection errors orderedContinuous	0.367177375735858
Number reset connection orderedContinuous	0.518986483707699
Number other errors orderedContinuous	0.552745712951224
Number disconnection errors orderedContinuous	0.552745712951224
Packet Destination IP nonKeyedSortedContinuous	0.51592596360935
Bytes Destination IP nonKeyedSortedContinuous	0.439296850206898
Connection Destination IP nonKeyedSortedContinuous	0.188811188811189
Packet Source IP nonKeyedSortedContinuous	0.36178106658408
Bytes Source IP nonKeyedSortedContinuous	0.207808665316558
Connection Source IP nonKeyedSortedContinuous	0.252744221904146

Table L.43: Similarity values for individual metrics of basecase 8,
train pair 3

Metric	Similarity
Packets in count	0.0107296137339056
Packets out count	0.00968523002421307
Connections in count	0.0114613180515759
Connections out count	0.00621118012422361
Bytes in count	0.00487770439827739
Bytes out count	0.0102618747032048
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0
Packet Service discrete	2.94976549364323e-05
Bytes Service discrete	1.14108774002275e-05
Connection Service discrete	6.82586091170744e-05
Continued on next page	

Table L.43 – continued from previous page

Metric	Similarity
Packet Source port discrete	0
Bytes Source port discrete	0
Connection Source port discrete	0
Connection Source port orderedContinuous	0
Packet TTL discrete	0.00255064762537363
Packet TTL orderedContinuous	0.0141956816144475
InterPacket delta sortedContinuous	0.4665067575339521
Packet sec orderedContinuous	0.0123886978198922
Packet min orderedContinuous	0.0164845566147066
Packet GmHour orderedContinuous	1
Packet LocHour orderedContinuous	1
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.0168541643659542
Bytes min orderedContinuous	0.0156395256863611
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.0180760170867322
Packets in last w secs orderedContinuous	0.01660208396094
Priv packets time rate sortedContinuous	0.8319226010955225
Unpriv packets time rate sortedContinuous	0.8892594178036315
Connections time rate sortedContinuous	0.3568361125575626
Priv connections connection time rate sortedContinuous	0.8888888888888888
Unpriv connections connection time rate sortedContinuous	0.8888888888888888
Priv packets priv connection time rate sortedContinuous	0.3337747320885113
Unpriv packets unpriv connection time rate sortedContinuous	0.8888888888888888
SYNs connection time rate sortedContinuous	0.6740740740740741
RSTs connection time rate sortedContinuous	0.7222222222222222
FINs connection time rate sortedContinuous	0.5555555555555556
PSH connection time rate sortedContinuous	0.4836266130383779
Establishment errors connection time rate sortedContinuous	0.8888888888888888
Continued on next page	

Table L.43 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.1292882080480194
Priv packets packet rate sortedContinuous	0.9046156651224097
Unpriv packets packet rate sortedContinuous	0.8904282826760102
InterConnection delta sortedContinuous	0.4218155701318225
Connection sec orderedContinuous	0.005623809056867
Connection min orderedContinuous	0.019762820789615
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.6304627608975433
Connection Priv connections rate sortedContinuous	0.9761904761904763
Connection Unpriv connections rate sortedContinuous	0.6666666666666666
Connection Priv packet rate sortedContinuous	0.6275050560764846
Connection Unpriv packet rate sortedContinuous	0.6666666666666666
Connection SYNs rate sortedContinuous	0.4999999999999999
Connection RSTs rate sortedContinuous	0.3333333333333333
Connection FINs rate sortedContinuous	0.3333333333333333
Connection PSH rate sortedContinuous	0.3333333333333333
Connection Establishment errors rate sortedContinuous	0.6666666666666666
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.3333333333333333
Number of packets orderedContinuous	0.0765651588563428
Number of packets in orderedContinuous	0.14348773950833
Number of packets out orderedContinuous	0.145130368832488
Duration sortedContinuous	0.6725062405946786
Number control packets rate sortedContinuous	0.9090909090909089
Number data packets rate sortedContinuous	0.6666666666666666
Number bytes transferred orderedContinuous	0.00341102698060109
Continued on next page	

Table L.43 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.0062094663593159
Number bytes transferred out orderedContinuous	0.010520175494345
Number data bytes transferred orderedContinuous	0.0123437589024895
Number data bytes transferred in orderedContinuous	0.0163623225680028
Number data bytes transferred out orderedContinuous	0.0269192177544275
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0
Max Dst Window orderedContinuous	0.00825082508250825
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	1
Wrong resend rate sortedContinuous	1
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	1
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.293807352365106
Number reset connection orderedContinuous	0.233703416168575
Number other errors orderedContinuous	0.354679802955665
Number disconnection errors orderedContinuous	0.354679802955665
Packet Destination IP nonKeyedSortedContinuous	0.338698140200286
Bytes Destination IP nonKeyedSortedContinuous	0.264977961412946
Connection Destination IP nonKeyedSortedContinuous	0.255730659025788
Packet Source IP nonKeyedSortedContinuous	0.206045949214027
Bytes Source IP nonKeyedSortedContinuous	0.100369032589698
Connection Source IP nonKeyedSortedContinuous	0.506191950464396

Table L.44: Similarity values for individual metrics of basecase 8,
test pair 1

Metric	Similarity
Packets in count	0.0149488388876002
Packets out count	0.0177227834436979
Connections in count	0.0578665367754505
Connections out count	0.0718288334182374
Bytes in count	0.00368189861262747
Bytes out count	0.0234031009119599
SYN-ONLY rate ratio	0.821110186615953
SYN-ACK rate ratio	0
Idle connection rate ratio	0.0391459074733095
Half-open connection rate ratio	0.869163673678811
Packet Service discrete	0.000168540877942803
Bytes Service discrete	0.000243273084910534
Connection Service discrete	0.000218215763799686
Packet Source port discrete	0.000172721444359097
Bytes Source port discrete	7.3420751558706e-05
Connection Source port discrete	0.000166967296303745
Connection Source port orderedContinuous	0.00121771234626171
Packet TTL discrete	0.048131045210037
Packet TTL orderedContinuous	0.104831311079864
InterPacket delta sortedContinuous	0.6122414477057327
Packet sec orderedContinuous	0.128656350813351
Packet min orderedContinuous	0.217096135349324
Packet GmHour orderedContinuous	0.12854763567979
Packet LocHour orderedContinuous	0.12854763567979
Packet weekday orderedContinuous	0.12854763567979
Bytes sec orderedContinuous	0.621273350990556
Bytes min orderedContinuous	0.289624028078725
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1

Continued on next page

Table L.44 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packets in last w secs orderedContinuous	0.00683376908504107
Priv packets time rate sortedContinuous	0.5906473109373714
Unpriv packets time rate sortedContinuous	0.7759684904181972
Connections time rate sortedContinuous	0.09212053490271885
Priv connections connection time rate sortedContinuous	0.6793578774203369
Unpriv connections connection time rate sortedContinuous	0.1219848293299621
Priv packets priv connection time rate sortedContinuous	0.1091407374374391
Unpriv packets unpriv connection time rate sortedContinuous	0.1036379635864534
SYNs connection time rate sortedContinuous	0.2954818305765547
RSTs connection time rate sortedContinuous	0.3346967513246671
FINs connection time rate sortedContinuous	0.2467291741739061
PSH connection time rate sortedContinuous	0.117953145786636
Establishment errors connection time rate sortedContinuous	0.2663834372985155
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.1807102567329438
Priv packets packet rate sortedContinuous	0.5944079886803236
Unpriv packets packet rate sortedContinuous	0.8206583570656584
InterConnection delta sortedContinuous	0.6665757991461602
Connection sec orderedContinuous	0.583238497500145
Connection min orderedContinuous	0.244488474877239
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.03243255006757962
Connection Priv connections rate sortedContinuous	0.848248933914939
Connection Unpriv connections rate sortedContinuous	0.07367207196445976
Connection Priv packet rate sortedContinuous	0.4286514158529137
Connection Unpriv packet rate sortedContinuous	0.03791132243727964
Connection SYNs rate sortedContinuous	0.4388528223330388
Continued on next page	

Table L.44 – continued from previous page

Metric	Similarity
Connection RSTs rate sortedContinuous	0.6922536641882572
Connection FINs rate sortedContinuous	0.8057391649720262
Connection PSH rate sortedContinuous	0.02105556319507156
Connection Establishment errors rate sortedContinuous	0.4118152218139515
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.08833625290493142
Number of packets orderedContinuous	0.0550210785434991
Number of packets in orderedContinuous	0.0645153314389768
Number of packets out orderedContinuous	0.0613824647139071
Duration sortedContinuous	0.1527919604633784
Number control packets rate sortedContinuous	0.7417209830987225
Number data packets rate sortedContinuous	0.3580490023425503
Number bytes transferred orderedContinuous	0.00202627740413617
Number bytes transferred in orderedContinuous	0.00375743160831549
Number bytes transferred out orderedContinuous	0.00421337245176325
Number data bytes transferred orderedContinuous	0.000660492303475872
Number data bytes transferred in orderedContinuous	0.00105521031624618
Number data bytes transferred out orderedContinuous	0.00183087957285612
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0307742699743024
Max Dst Window orderedContinuous	0.00328692333369995
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9597268016957137
Wrong resend rate sortedContinuous	0.9268547544409614
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9545454545454546
Window exceeded rate sortedContinuous	0.9545454545454546
Hole rate sortedContinuous	0.9714587737843552
Continued on next page	

Table L.44 – continued from previous page

Metric	Similarity
Number connection errors orderedContinuous	0.0933200909027718
Packet Destination IP nonKeyedSortedContinuous	0.540105105511022
Bytes Destination IP nonKeyedSortedContinuous	0.49834140805827
Connection Destination IP nonKeyedSortedContinuous	0.541931465899026
Packet Source IP nonKeyedSortedContinuous	0.283008374186569
Bytes Source IP nonKeyedSortedContinuous	0.198785044969257
Connection Source IP nonKeyedSortedContinuous	0.134743613408359

Table L.45: Similarity values for individual metrics of basecase 8,
test pair 2

Metric	Similarity
Packets in count	0.000846955657377801
Packets out count	0.000894469719485147
Connections in count	0.00186811133943587
Connections out count	0.000914494741655281
Bytes in count	0.000567045440442282
Bytes out count	0.000755926643515648
SYN-ONLY rate ratio	0.122522522522523
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0.501438159156281
Packet Service discrete	0.000286000373683938
Bytes Service discrete	0.000209872754634503
Connection Service discrete	0.000113333435224312
Packet Source port discrete	0.000229968972887416
Bytes Source port discrete	0.000185077087025348
Connection Source port discrete	7.05371054510262e-06
Connection Source port orderedContinuous	1.69913848399447e-05
Continued on next page	

Table L.45 – continued from previous page

Metric	Similarity
Packet TTL discrete	0.0257569820271604
Packet TTL orderedContinuous	0.0882814229430758
InterPacket delta sortedContinuous	0.04475384794280565
Packet sec orderedContinuous	0.399922883483774
Packet min orderedContinuous	0.29935852859227
Packet GmHour orderedContinuous	0.960680284082502
Packet LocHour orderedContinuous	0.978259868583948
Packet weekday orderedContinuous	0.980865799141566
Bytes sec orderedContinuous	0.380038672904581
Bytes min orderedContinuous	0.3010530243847
Bytes GmHour orderedContinuous	0.944838816226003
Bytes LocHour orderedContinuous	0.969725356144982
Bytes weekday orderedContinuous	0.972672270882412
Packet size orderedContinuous	0.0296106758401483
Packets in last w secs orderedContinuous	4.89852035621413e-07
Priv packets time rate sortedContinuous	0.9681105420159451
Unpriv packets time rate sortedContinuous	0.9672751079919086
Connections time rate sortedContinuous	0.254508767102197
Priv connections connection time rate sortedContinuous	0.8490740719939246
Unpriv connections connection time rate sortedContinuous	0.005020145998252265
Priv packets priv connection time rate sortedContinuous	0.2076558755500037
Unpriv packets unpriv connection time rate sortedContinuous	0.002283804099168922
SYNs connection time rate sortedContinuous	0.1635150960918547
RSTs connection time rate sortedContinuous	0.08722296732139252
FINs connection time rate sortedContinuous	0.03127045271155405
PSH connection time rate sortedContinuous	0.3221779825643865
Establishment errors connection time rate sortedContinuous	0.02724358974358974
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9519230769230769
Ave duration over last w secs sortedContinuous	0.01061205485852535
Priv packets packet rate sortedContinuous	0.9666789296423395
Continued on next page	

Table L.45 – continued from previous page

Metric	Similarity
Unpriv packets packet rate sortedContinuous	0.9754739654133325
InterConnection delta sortedContinuous	0.05926133458742554
Connection sec orderedContinuous	0.562061523684802
Connection min orderedContinuous	0.339253449982436
Connection GmHour orderedContinuous	0.689609676657704
Connection LocHour orderedContinuous	0.827609327553982
Connection weekday orderedContinuous	0.856953282459907
Connection packet rate sortedContinuous	0.2119816383210661
Connection Priv connections rate sortedContinuous	0.9577434239339408
Connection Unpriv connections rate sortedContinuous	0.7641613098409387
Connection Priv packet rate sortedContinuous	0.2084752519992381
Connection Unpriv packet rate sortedContinuous	0.5810352627301699
Connection SYNs rate sortedContinuous	0.4514847179805689
Connection RSTs rate sortedContinuous	0.2547430253347184
Connection FINs rate sortedContinuous	0.4737483075244059
Connection PSH rate sortedContinuous	0.2240328307267655
Connection Establishment errors rate sortedContinuous	0.6923076923076923
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9743589743589743
Ave duration over last m connections sortedContinuous	0.2581362244725532
Number of packets orderedContinuous	0.00903580485496807
Number of packets in orderedContinuous	0.0164271331177056
Number of packets out orderedContinuous	0.0159254140067114
Duration sortedContinuous	0.9000531230029042
Number control packets rate sortedContinuous	0.8587554766839047
Number data packets rate sortedContinuous	0.8848736733352118
Number bytes transferred orderedContinuous	0.00298073951889832
Number bytes transferred in orderedContinuous	0.00128968262423323
Number bytes transferred out orderedContinuous	0.00259184709306816
Number data bytes transferred orderedContinuous	0.000429011208226235
Number data bytes transferred in orderedContinuous	0.000586752060369876
Continued on next page	

Table L.45 – continued from previous page

Metric	Similarity
Number data bytes transferred out orderedContinuous	0.000814682023550269
Fragmented packets rate sortedContinuous	0.9743589743589743
Bad fragment rate sortedContinuous	0.9743589743589743
Max Src Window orderedContinuous	0.00297934680796712
Max Dst Window orderedContinuous	0.000124258963958403
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.875
Wrong resend rate sortedContinuous	0.875
Duplicate ACK rate sortedContinuous	0.875
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.875
Hole rate sortedContinuous	0.875
Number connection errors orderedContinuous	0.0703389677024493
Number reset connection orderedContinuous	0.426425954623344
Number other errors orderedContinuous	0.419910929863295
Number disconnection errors orderedContinuous	0.228968960267532
Packet Destination IP nonKeyedSortedContinuous	0.487644021818451
Bytes Destination IP nonKeyedSortedContinuous	0.335247815348047
Connection Destination IP nonKeyedSortedContinuous	0.0690592401186716
Packet Source IP nonKeyedSortedContinuous	0.568601021110763
Bytes Source IP nonKeyedSortedContinuous	0.432913338906541
Connection Source IP nonKeyedSortedContinuous	0.54776129850959

Table L.46: Similarity values for individual metrics of basecase 8,
test pair 3

Metric	Similarity
Packets in count	0.305301645338208
Continued on next page	

Table L.46 – continued from previous page

Metric	Similarity
Packets out count	0.0681551116333725
Connections in count	0.647173489278752
Connections out count	0.165714285714286
Bytes in count	0.500069963474241
Bytes out count	0.285166699940108
SYN-ONLY rate ratio	0.2222222222222222
SYN-ACK rate ratio	1
Idle connection rate ratio	1
Half-open connection rate ratio	0.2222222222222222
Packet Service discrete	0.00509656010048056
Bytes Service discrete	0.00059218937988254
Connection Service discrete	0.00171421801866588
Packet Source port discrete	1.51408881912207e-05
Bytes Source port discrete	0.000421698388846131
Connection Source port discrete	0.000566233296170741
Connection Source port orderedContinuous	0.000997393562765208
Packet TTL discrete	0.0109323957319429
Packet TTL orderedContinuous	0.0236791157046855
InterPacket delta sortedContinuous	0.09193294110755039
Packet sec orderedContinuous	0.0290877094771337
Packet min orderedContinuous	0.0423573670292664
Packet GmHour orderedContinuous	0.0192525481313702
Packet LocHour orderedContinuous	0.0192525481313702
Packet weekday orderedContinuous	0
Bytes sec orderedContinuous	0.669499035620405
Bytes min orderedContinuous	0.542420240294305
Bytes GmHour orderedContinuous	1
Bytes LocHour orderedContinuous	1
Bytes weekday orderedContinuous	0
Packet size orderedContinuous	0
Packets in last w secs orderedContinuous	0.00745302057701716
Continued on next page	

Table L.46 – continued from previous page

Metric	Similarity
Priv packets time rate sortedContinuous	0.3097689745396582
Unpriv packets time rate sortedContinuous	0.5407677901195822
Connections time rate sortedContinuous	0.2298103336529433
Priv connections connection time rate sortedContinuous	0.2452542921930677
Unpriv connections connection time rate sortedContinuous	0.454421768707483
Priv packets priv connection time rate sortedContinuous	0.2340495757554075
Unpriv packets unpriv connection time rate sortedContinuous	0.4526643990929705
SYNs connection time rate sortedContinuous	0.03217027942661902
RSTs connection time rate sortedContinuous	0.7397959183673469
FINs connection time rate sortedContinuous	0.5102040816326531
PSH connection time rate sortedContinuous	0.1377551020408163
Establishment errors connection time rate sortedContinuous	0.973469387755102
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.05623454153115619
Priv packets packet rate sortedContinuous	0.477859942845179
Unpriv packets packet rate sortedContinuous	0.8914408915774519
InterConnection delta sortedContinuous	0.1838468922792464
Connection sec orderedContinuous	0.66318081060017
Connection min orderedContinuous	0.559045876929362
Connection GmHour orderedContinuous	1
Connection LocHour orderedContinuous	1
Connection weekday orderedContinuous	0
Connection packet rate sortedContinuous	0.6276241910335594
Connection Priv connections rate sortedContinuous	0.3198558812156493
Connection Unpriv connections rate sortedContinuous	0.06568978202072481
Connection Priv packet rate sortedContinuous	0.6139866031767193
Connection Unpriv packet rate sortedContinuous	0.2978698626309963
Connection SYNs rate sortedContinuous	0.16327145533136
Connection RSTs rate sortedContinuous	0.3435897435897436
Connection FINs rate sortedContinuous	0.1025641025641026

Continued on next page

Table L.46 – continued from previous page

Metric	Similarity
Connection PSH rate sortedContinuous	0.04102564102564103
Connection Establishment errors rate sortedContinuous	0.8746438746438747
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.007908017825557759
Number of packets orderedContinuous	0.0228439370685286
Number of packets in orderedContinuous	0.036354115542091
Number of packets out orderedContinuous	0.00627684823200574
Duration sortedContinuous	0.1438158589796473
Number control packets rate sortedContinuous	0.06
Number data packets rate sortedContinuous	0.5986072665310046
Number bytes transferred orderedContinuous	0.00426084659648167
Number bytes transferred in orderedContinuous	0.00399719634071934
Number bytes transferred out orderedContinuous	0.00120145354639952
Number data bytes transferred orderedContinuous	0.000402079864058059
Number data bytes transferred in orderedContinuous	0.000620983651752369
Number data bytes transferred out orderedContinuous	0.000775290883175593
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0243271221532091
Max Dst Window orderedContinuous	0.0198019801980198
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.6666666666666666
Wrong resend rate sortedContinuous	0.6666666666666666
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.6666666666666666
Window exceeded rate sortedContinuous	0.6666666666666666
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.393214037431377
Number reset connection orderedContinuous	0
Continued on next page	

Table L.46 – continued from previous page

Metric	Similarity
Number other errors orderedContinuous	0
Number disconnection errors orderedContinuous	0
Packet Destination IP nonKeyedSortedContinuous	0.535371551156096
Bytes Destination IP nonKeyedSortedContinuous	0.440158618504229
Connection Destination IP nonKeyedSortedContinuous	0.515724496426251
Packet Source IP nonKeyedSortedContinuous	0.347756518644087
Bytes Source IP nonKeyedSortedContinuous	0.634899475377478
Connection Source IP nonKeyedSortedContinuous	0.776808957004962

Table L.47: Similarity values for individual metrics of basecase 9,
train pair 1

Metric	Similarity
Packets in count	0.723002868673117
Packets out count	0.67972082202404
Connections in count	0.950193088424596
Connections out count	0.951135876553793
Bytes in count	0.947419067121155
Bytes out count	0.21151812670363
SYN-ONLY rate ratio	0.956183898122166
SYN-ACK rate ratio	0.567887931034482
Idle connection rate ratio	1
Half-open connection rate ratio	0.982261640798226
Packet Service discrete	0.00345466287215414
Bytes Service discrete	0.00370629432524323
Connection Service discrete	0.00379362628434976
Packet Source port discrete	0.00237757437151189
Bytes Source port discrete	0.00182826551953173
Connection Source port discrete	0.0033553038193404
Continued on next page	

Table L.47 – continued from previous page

Metric	Similarity
Connection Source port orderedContinuous	0.0257451199831019
Packet TTL discrete	0.355390727401179
Packet TTL orderedContinuous	0.408040766782697
InterPacket delta sortedContinuous	0.69846218626056
Packet sec orderedContinuous	0.80021578112537
Packet min orderedContinuous	0.713591264377809
Packet GmHour orderedContinuous	0.681876723850178
Packet LocHour orderedContinuous	0.664464569673679
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.566981950954468
Bytes min orderedContinuous	0.646002642684187
Bytes GmHour orderedContinuous	0.699524389224277
Bytes LocHour orderedContinuous	0.675085715814845
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.442844786089476
Packets in last w secs orderedContinuous	0.635676591922927
Priv packets time rate sortedContinuous	0.963409350808187
Unpriv packets time rate sortedContinuous	0.9718071630372072
Connections time rate sortedContinuous	0.5607969980375018
Priv connections connection time rate sortedContinuous	0.9945891124810067
Unpriv connections connection time rate sortedContinuous	0.985357998468639
Priv packets priv connection time rate sortedContinuous	0.5621428905854481
Unpriv packets unpriv connection time rate sortedContinuous	0.9832803275157578
SYNs connection time rate sortedContinuous	0.4803232123109374
RSTs connection time rate sortedContinuous	0.8690783337296844
FINs connection time rate sortedContinuous	0.7769101127350391
PSH connection time rate sortedContinuous	0.3589460431548168
Establishment errors connection time rate sortedContinuous	0.9892034431261059
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.2547246416055606
Continued on next page	

Table L.47 – continued from previous page

Metric	Similarity
Priv packets packet rate sortedContinuous	0.9655528550676811
Unpriv packets packet rate sortedContinuous	0.977397835199547
InterConnection delta sortedContinuous	0.924074756658263
Connection sec orderedContinuous	0.841483785757011
Connection min orderedContinuous	0.847431250649707
Connection GmHour orderedContinuous	0.893638568953177
Connection LocHour orderedContinuous	0.89091574658126
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.8128436996423166
Connection Priv connections rate sortedContinuous	0.9976696585912743
Connection Unpriv connections rate sortedContinuous	0.9194819209469525
Connection Priv packet rate sortedContinuous	0.811239727133618
Connection Unpriv packet rate sortedContinuous	0.9195944650906378
Connection SYNs rate sortedContinuous	0.9401616070967103
Connection RSTs rate sortedContinuous	0.9776214272915709
Connection FINs rate sortedContinuous	0.8992785398373478
Connection PSH rate sortedContinuous	0.7311136963901049
Connection Establishment errors rate sortedContinuous	0.9500206337941647
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.7006644015627381
Number of packets orderedContinuous	0.401186077987924
Number of packets in orderedContinuous	0.430156040586342
Number of packets out orderedContinuous	0.421953913889703
Duration sortedContinuous	0.9806479298707476
Number control packets rate sortedContinuous	0.9940008963384308
Number data packets rate sortedContinuous	0.9949499144763842
Number bytes transferred orderedContinuous	0.169133525867807
Number bytes transferred in orderedContinuous	0.206050546355047
Number bytes transferred out orderedContinuous	0.221809131055928
Number data bytes transferred orderedContinuous	0.0855223049790045
Continued on next page	

Table L.47 – continued from previous page

Metric	Similarity
Number data bytes transferred in orderedContinuous	0.097616241633632
Number data bytes transferred out orderedContinuous	0.117619223213011
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.209580781930074
Max Dst Window orderedContinuous	0.147471038408643
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9408962362709255
Wrong resend rate sortedContinuous	0.9949615153857591
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9974201870364398
Window exceeded rate sortedContinuous	0.9765023806152822
Hole rate sortedContinuous	0.991056426669346
Number connection errors orderedContinuous	0.798334721164517
Number reset connection orderedContinuous	0.96840686443213
Number other errors orderedContinuous	0.962596324691699
Number disconnection errors orderedContinuous	0.962596324691699
Packet Destination IP nonKeyedSortedContinuous	0.937500236455141
Bytes Destination IP nonKeyedSortedContinuous	0.946058538197001
Connection Destination IP nonKeyedSortedContinuous	0.838077293783133
Packet Source IP nonKeyedSortedContinuous	0.917159832588989
Bytes Source IP nonKeyedSortedContinuous	0.957617280008284
Connection Source IP nonKeyedSortedContinuous	0.976407450151679

Table L.48: Similarity values for individual metrics of basecase 9,
train pair 2

Metric	Similarity
Packets in count	0.247379295763529
Packets out count	0.233965993987594
Connections in count	0.546070844343778
Connections out count	0.377393421602386
Bytes in count	0.189021722554837
Bytes out count	0.0873973393502604
SYN-ONLY rate ratio	0.83145887538425
SYN-ACK rate ratio	0.464487595118405
Idle connection rate ratio	0.31477147527891
Half-open connection rate ratio	0.827165664577193
Packet Service discrete	0.107468066438766
Bytes Service discrete	0.0873332322141899
Connection Service discrete	0.0755027239805489
Packet Source port discrete	0.1070526934799
Bytes Source port discrete	0.0853444944593344
Connection Source port discrete	0.157307232939378
Connection Source port orderedContinuous	0.160712575004706
Packet TTL discrete	0.197699907791339
Packet TTL orderedContinuous	0.245654516313054
InterPacket delta sortedContinuous	0.4511234457479123
Packet sec orderedContinuous	0.975065391513986
Packet min orderedContinuous	0.910918184787184
Packet GmHour orderedContinuous	0.757857742704848
Packet LocHour orderedContinuous	0.780627267728397
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.959081903906716
Bytes min orderedContinuous	0.877050635284214
Bytes GmHour orderedContinuous	0.677700187266954
Bytes LocHour orderedContinuous	0.681810224718891
Continued on next page	

Table L.48 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.44221141075243
Packets in last w secs orderedContinuous	0.209750958281936
Priv packets time rate sortedContinuous	0.1785191654332786
Unpriv packets time rate sortedContinuous	0.8375964820333232
Connections time rate sortedContinuous	0.4128587964530602
Priv connections connection time rate sortedContinuous	0.6493418828042411
Unpriv connections connection time rate sortedContinuous	0.6675367378012951
Priv packets priv connection time rate sortedContinuous	0.7548570184130979
Unpriv packets unpriv connection time rate sortedContinuous	0.5184137183324557
SYNs connection time rate sortedContinuous	0.7246636675337225
RSTs connection time rate sortedContinuous	0.7235034509698217
FINs connection time rate sortedContinuous	0.7690848743337452
PSH connection time rate sortedContinuous	0.4300396082321281
Establishment errors connection time rate sortedContinuous	0.7737930099805811
Other errors connection time rate sortedContinuous	0.9987494654660327
Disconnection errors connection time rate sortedContinuous	0.9965843212989018
Ave duration over last w secs sortedContinuous	0.7423297860425393
Priv packets packet rate sortedContinuous	0.3447010003251225
Unpriv packets packet rate sortedContinuous	0.8320017147440516
InterConnection delta sortedContinuous	0.4380161294250676
Connection sec orderedContinuous	0.977986662073211
Connection min orderedContinuous	0.747808720886706
Connection GmHour orderedContinuous	0.587518351728927
Connection LocHour orderedContinuous	0.620754792913452
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.7267280646836572
Connection Priv connections rate sortedContinuous	0.6045250419349643
Connection Unpriv connections rate sortedContinuous	0.2130399577886477
Connection Priv packet rate sortedContinuous	0.7401054946185216
Connection Unpriv packet rate sortedContinuous	0.1107974832588941
Continued on next page	

Table L.48 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.7619992949911174
Connection RSTs rate sortedContinuous	0.739206986733156
Connection FINs rate sortedContinuous	0.7253849689124724
Connection PSH rate sortedContinuous	0.4748280012196687
Connection Establishment errors rate sortedContinuous	0.8468581222961757
Connection Other errors rate sortedContinuous	0.9999619924160251
Connection Disconnection errors rate sortedContinuous	0.9984239985918167
Ave duration over last m connections sortedContinuous	0.55731420889861
Number of packets orderedContinuous	0.159399443536085
Number of packets in orderedContinuous	0.157227067250452
Number of packets out orderedContinuous	0.15867625938959
Duration sortedContinuous	0.9231384065269166
Number control packets rate sortedContinuous	0.9450382618547318
Number data packets rate sortedContinuous	0.9420579720478777
Number bytes transferred orderedContinuous	0.137071713292591
Number bytes transferred in orderedContinuous	0.126429059757932
Number bytes transferred out orderedContinuous	0.179542104557986
Number data bytes transferred orderedContinuous	0.125088483586874
Number data bytes transferred in orderedContinuous	0.119963262916454
Number data bytes transferred out orderedContinuous	0.152862832781754
Fragmented packets rate sortedContinuous	0.9999469086061149
Bad fragment rate sortedContinuous	0.9999469086061149
Max Src Window orderedContinuous	0.109633796699537
Max Dst Window orderedContinuous	0.178265417165691
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9721289928663975
Wrong resend rate sortedContinuous	0.9728995922892119
Duplicate ACK rate sortedContinuous	0.9826678253516471
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9449120648730737
Window exceeded rate sortedContinuous	0.9960812382016851
Continued on next page	

Table L.48 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9787559842425472
Number connection errors orderedContinuous	0.258198966015983
Number reset connection orderedContinuous	0.821674856737364
Number other errors orderedContinuous	0.739493608570879
Number disconnection errors orderedContinuous	0.458838946044496
Packet Destination IP nonKeyedSortedContinuous	0.804966325343872
Bytes Destination IP nonKeyedSortedContinuous	0.804242571981551
Connection Destination IP nonKeyedSortedContinuous	0.871190295105794
Packet Source IP nonKeyedSortedContinuous	0.364713500376237
Bytes Source IP nonKeyedSortedContinuous	0.174099853557884
Connection Source IP nonKeyedSortedContinuous	0.200416187484891

Table L.49: Similarity values for individual metrics of basecase 9,
train pair 3

Metric	Similarity
Packets in count	0.00927106269556144
Packets out count	0.00806180718844474
Connections in count	0.045611610228058
Connections out count	0.298850574712644
Bytes in count	0.00149018514836452
Bytes out count	0.0180519462143972
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	0.000466329471954043
Bytes Service discrete	0.00082535500753957
Connection Service discrete	0.00890608868186299
Continued on next page	

Table L.49 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.000260980464302827
Bytes Source port discrete	0.000251898852199953
Connection Source port discrete	0.000548033113050648
Connection Source port orderedContinuous	0.00244688782338049
Packet TTL discrete	0.0599189775518195
Packet TTL orderedContinuous	0.161982054483062
InterPacket delta sortedContinuous	0.05447611212254226
Packet sec orderedContinuous	0.645977279818682
Packet min orderedContinuous	0.376898077666883
Packet GmHour orderedContinuous	0.371110778119728
Packet LocHour orderedContinuous	0.387677156552819
Packet weekday orderedContinuous	0.25
Bytes sec orderedContinuous	0.644296817293207
Bytes min orderedContinuous	0.212848932063895
Bytes GmHour orderedContinuous	0.248436163762692
Bytes LocHour orderedContinuous	0.291502558934066
Bytes weekday orderedContinuous	0.25
Packet size orderedContinuous	0.00595516713542601
Packets in last w secs orderedContinuous	0.00034642363155271
Priv packets time rate sortedContinuous	0.7770285228673961
Unpriv packets time rate sortedContinuous	0.745513388147512
Connections time rate sortedContinuous	0.2162456449116893
Priv connections connection time rate sortedContinuous	0.978401266323704
Unpriv connections connection time rate sortedContinuous	0.9766941391941391
Priv packets priv connection time rate sortedContinuous	0.2697088982241802
Unpriv packets unpriv connection time rate sortedContinuous	0.9440627098935864
SYNs connection time rate sortedContinuous	0.63638197785854
RSTs connection time rate sortedContinuous	0.5491530525058823
FINs connection time rate sortedContinuous	0.5326223904476913
PSH connection time rate sortedContinuous	0.2799083978298966
Establishment errors connection time rate sortedContinuous	0.7815090701571395
Continued on next page	

Table L.49 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	0.9928571428571429
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.03322757979202155
Priv packets packet rate sortedContinuous	0.7656872490084568
Unpriv packets packet rate sortedContinuous	0.8722690813505699
InterConnection delta sortedContinuous	0.2481771567282734
Connection sec orderedContinuous	0.594502070777586
Connection min orderedContinuous	0.418846796282293
Connection GmHour orderedContinuous	0.320768920930097
Connection LocHour orderedContinuous	0.329009718521742
Connection weekday orderedContinuous	0.25
Connection packet rate sortedContinuous	0.6059343252107818
Connection Priv connections rate sortedContinuous	0.9504151393163392
Connection Unpriv connections rate sortedContinuous	0.4201986909261881
Connection Priv packet rate sortedContinuous	0.6302384853496599
Connection Unpriv packet rate sortedContinuous	0.4206153545607732
Connection SYNs rate sortedContinuous	0.5172697038864708
Connection RSTs rate sortedContinuous	0.3420001066683081
Connection FINs rate sortedContinuous	0.1524435122303671
Connection PSH rate sortedContinuous	0.7411468593243449
Connection Establishment errors rate sortedContinuous	0.3773838220204827
Connection Other errors rate sortedContinuous	0.9782608695652174
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.2246370530267475
Number of packets orderedContinuous	0.0512884721352545
Number of packets in orderedContinuous	0.0606056376356697
Number of packets out orderedContinuous	0.0653195310313783
Duration sortedContinuous	0.07443016840869854
Number control packets rate sortedContinuous	0.2973866811455327
Number data packets rate sortedContinuous	0.232284654977487
Number bytes transferred orderedContinuous	0.00544057096868665

Continued on next page

Table L.49 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.00576120747783369
Number bytes transferred out orderedContinuous	0.00601263518731557
Number data bytes transferred orderedContinuous	0.0053129668665828
Number data bytes transferred in orderedContinuous	0.0051169025691711
Number data bytes transferred out orderedContinuous	0.0111521561871032
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0929618551752151
Max Dst Window orderedContinuous	0.0246570055044356
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.6417758019944732
Wrong resend rate sortedContinuous	0.9935064935064935
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9285714285714286
Window exceeded rate sortedContinuous	0.9285714285714286
Hole rate sortedContinuous	0.9285714285714286
Number connection errors orderedContinuous	0.135220260417709
Number reset connection orderedContinuous	0.542063013941325
Number other errors orderedContinuous	0.551062324719357
Number disconnection errors orderedContinuous	0.48464544946957
Packet Destination IP nonKeyedSortedContinuous	0.653903801295796
Bytes Destination IP nonKeyedSortedContinuous	0.620324875933345
Connection Destination IP nonKeyedSortedContinuous	0.827067495968671
Packet Source IP nonKeyedSortedContinuous	0.753182307158014
Bytes Source IP nonKeyedSortedContinuous	0.662522647289961
Connection Source IP nonKeyedSortedContinuous	0.853338665656332

Table L.50: Similarity values for individual metrics of basecase 9,
test pair 1

Metric	Similarity
Packets in count	0.330169449807321
Packets out count	0.394381339690553
Connections in count	0.909330175023029
Connections out count	0.757924641148325
Bytes in count	0.204937799628595
Bytes out count	0.457329796453125
SYN-ONLY rate ratio	0.395402260091801
SYN-ACK rate ratio	1
Idle connection rate ratio	0.688430871069655
Half-open connection rate ratio	0.60058265510594
Packet Service discrete	0.0444460759564044
Bytes Service discrete	0.0315778340115401
Connection Service discrete	0.0348127889600975
Packet Source port discrete	0.0372495954631472
Bytes Source port discrete	0.0335998000086102
Connection Source port discrete	0.0710692691473405
Connection Source port orderedContinuous	0.0816210210327635
Packet TTL discrete	0.312893286551698
Packet TTL orderedContinuous	0.383681291314602
InterPacket delta sortedContinuous	0.6918298313984448
Packet sec orderedContinuous	0.767023951308846
Packet min orderedContinuous	0.610486278871729
Packet GmHour orderedContinuous	0.44694859514471
Packet LocHour orderedContinuous	0.471590292337965
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.842863311330037
Bytes min orderedContinuous	0.48578195326172
Bytes GmHour orderedContinuous	0.344571352404742
Bytes LocHour orderedContinuous	0.361497180378317
Continued on next page	

Table L.50 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.570219928839189
Packets in last w secs orderedContinuous	0.119978380678036
Priv packets time rate sortedContinuous	0.910985346911116
Unpriv packets time rate sortedContinuous	0.9269071167929702
Connections time rate sortedContinuous	0.2351566709502367
Priv connections connection time rate sortedContinuous	0.9426377080324039
Unpriv connections connection time rate sortedContinuous	0.7881331099881217
Priv packets priv connection time rate sortedContinuous	0.2958741787890561
Unpriv packets unpriv connection time rate sortedContinuous	0.5617358013529395
SYNs connection time rate sortedContinuous	0.5539421133816953
RSTs connection time rate sortedContinuous	0.9313418393826089
FINs connection time rate sortedContinuous	0.8130998960837432
PSH connection time rate sortedContinuous	0.3179925911063577
Establishment errors connection time rate sortedContinuous	0.8061765356306474
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9874629670803184
Ave duration over last w secs sortedContinuous	0.1153502474258399
Priv packets packet rate sortedContinuous	0.9391283293846563
Unpriv packets packet rate sortedContinuous	0.9417338487826956
InterConnection delta sortedContinuous	0.9416289639082522
Connection sec orderedContinuous	0.763336499907071
Connection min orderedContinuous	0.89341571692288
Connection GmHour orderedContinuous	0.689421386228485
Connection LocHour orderedContinuous	0.714845646183531
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.8432282571830971
Connection Priv connections rate sortedContinuous	0.9476802836793148
Connection Unpriv connections rate sortedContinuous	0.8073143351904563
Connection Priv packet rate sortedContinuous	0.8741345615836421
Connection Unpriv packet rate sortedContinuous	0.9110758194526553
Continued on next page	

Table L.50 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.8474834977908246
Connection RSTs rate sortedContinuous	0.8129490426829971
Connection FINs rate sortedContinuous	0.8381916845773695
Connection PSH rate sortedContinuous	0.7778528340159025
Connection Establishment errors rate sortedContinuous	0.7088440597281267
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9940840916047528
Ave duration over last m connections sortedContinuous	0.5499024724155378
Number of packets orderedContinuous	0.335445064565001
Number of packets in orderedContinuous	0.360629318700637
Number of packets out orderedContinuous	0.341664555474272
Duration sortedContinuous	0.813722818315873
Number control packets rate sortedContinuous	0.9181495739319351
Number data packets rate sortedContinuous	0.9548124315402088
Number bytes transferred orderedContinuous	0.104286956146951
Number bytes transferred in orderedContinuous	0.106465778015199
Number bytes transferred out orderedContinuous	0.220774068478903
Number data bytes transferred orderedContinuous	0.0886031205276551
Number data bytes transferred in orderedContinuous	0.0914164094691252
Number data bytes transferred out orderedContinuous	0.18791775970171
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.391243205032492
Max Dst Window orderedContinuous	0.115472031327091
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.7733282514672691
Wrong resend rate sortedContinuous	0.983010050924241
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.8593148402038685
Continued on next page	

Table L.50 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9991683295940659
Number connection errors orderedContinuous	0.252751736134866
Number reset connection orderedContinuous	0.822100805234976
Number other errors orderedContinuous	0.842025045156489
Number disconnection errors orderedContinuous	0.841908024435963
Packet Destination IP nonKeyedSortedContinuous	0.907053455320735
Bytes Destination IP nonKeyedSortedContinuous	0.874269113786691
Connection Destination IP nonKeyedSortedContinuous	0.91926500232295
Packet Source IP nonKeyedSortedContinuous	0.956417391308371
Bytes Source IP nonKeyedSortedContinuous	0.943590065705634
Connection Source IP nonKeyedSortedContinuous	0.981358601800174

Table L.51: Similarity values for individual metrics of basecase 9,
test pair 2

Metric	Similarity
Packets in count	0.727950271483096
Packets out count	0.912399882924893
Connections in count	0.847227841407564
Connections out count	0.948052964569879
Bytes in count	0.559377268813724
Bytes out count	0.984270797125653
SYN-ONLY rate ratio	0.976423369730494
SYN-ACK rate ratio	0.569980419704741
Idle connection rate ratio	0.764762114495962
Half-open connection rate ratio	0.981322496452911
Packet Service discrete	0.253370671676649
Bytes Service discrete	0.21010958642174
Connection Service discrete	0.207617826637762
Continued on next page	

Table L.51 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.368418516117819
Bytes Source port discrete	0.316994515104072
Connection Source port discrete	0.394935614070193
Connection Source port orderedContinuous	0.379618230530831
Packet TTL discrete	0.551023281594491
Packet TTL orderedContinuous	0.656287690516908
InterPacket delta sortedContinuous	0.9236050415563166
Packet sec orderedContinuous	0.991583839463241
Packet min orderedContinuous	0.96892097244854
Packet GmHour orderedContinuous	0.641293086442929
Packet LocHour orderedContinuous	0.639230740300303
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.975902680408007
Bytes min orderedContinuous	0.904084663386956
Bytes GmHour orderedContinuous	0.753417671779448
Bytes LocHour orderedContinuous	0.732316934727167
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.434780594338045
Packets in last w secs orderedContinuous	0.608220188874308
Priv packets time rate sortedContinuous	0.6110366836392053

Table L.52: Similarity values for individual metrics of basecase 9, test pair 3

Metric	Similarity
Packets in count	0.917713718940426
Packets out count	0.983065504613372
Connections in count	0.952981301829877
Connections out count	0.978973279160117
Continued on next page	

Table L.52 – continued from previous page

Metric	Similarity
Bytes in count	0.860355896158552
Bytes out count	0.983375411795266
SYN-ONLY rate ratio	0.567053101454186
SYN-ACK rate ratio	0.680089867605799
Idle connection rate ratio	0.791168663504131
Half-open connection rate ratio	0.802324777329111
Packet Service discrete	0.151217719279793
Bytes Service discrete	0.111307298336943
Connection Service discrete	0.0806214235113318
Packet Source port discrete	0.225103264335135
Bytes Source port discrete	0.192134278469955
Connection Source port discrete	0.314798158609293
Connection Source port orderedContinuous	0.331035577329619
Packet TTL discrete	0.464700900590507
Packet TTL orderedContinuous	0.504756496625475
InterPacket delta sortedContinuous	0.96915599628322
Packet sec orderedContinuous	0.989649077967776
Packet min orderedContinuous	0.864739505812799
Packet GmHour orderedContinuous	0.588242064968561
Packet LocHour orderedContinuous	0.594072215470734
Packet weekday orderedContinuous	0.220618609292495
Bytes sec orderedContinuous	0.984405301589555
Bytes min orderedContinuous	0.832185789963252
Bytes GmHour orderedContinuous	0.56703505847358
Bytes LocHour orderedContinuous	0.5805280832909
Bytes weekday orderedContinuous	0.220920181989876
Packet size orderedContinuous	0.718900073224703
Packets in last w secs orderedContinuous	0.439787140074172
Priv packets time rate sortedContinuous	0.9146927004621219
Unpriv packets time rate sortedContinuous	0.9640875328287065
Connections time rate sortedContinuous	0.8493669265901259
Continued on next page	

Table L.52 – continued from previous page

Metric	Similarity
Priv connections connection time rate sortedContinuous	0.8574329834450748
Unpriv connections connection time rate sortedContinuous	0.7078925137332508
Priv packets priv connection time rate sortedContinuous	0.9002767306267965
Unpriv packets unpriv connection time rate sortedContinuous	0.6205723512492643
SYNs connection time rate sortedContinuous	0.9195314521493504
RSTs connection time rate sortedContinuous	0.9435689333816857
FINs connection time rate sortedContinuous	0.8880039899848294
PSH connection time rate sortedContinuous	0.897096076366942
Establishment errors connection time rate sortedContinuous	0.5566655247533859
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9388741349529784
Ave duration over last w secs sortedContinuous	0.8827499584635822
Priv packets packet rate sortedContinuous	0.9023362418320894
Unpriv packets packet rate sortedContinuous	0.9655749689234561
InterConnection delta sortedContinuous	0.9407979702291307
Connection sec orderedContinuous	0.981283022329277
Connection min orderedContinuous	0.864218762519744
Connection GmHour orderedContinuous	0.557339764686966
Connection LocHour orderedContinuous	0.56475610344554
Connection weekday orderedContinuous	0.217638091841669
Connection packet rate sortedContinuous	0.8306604158864706
Connection Priv connections rate sortedContinuous	0.9848699284152394
Connection Unpriv connections rate sortedContinuous	0.8354933256490774
Connection Priv packet rate sortedContinuous	0.773687704825451
Connection Unpriv packet rate sortedContinuous	0.4996474271266507
Connection SYNs rate sortedContinuous	0.9311750954610741
Connection RSTs rate sortedContinuous	0.9857473952617979
Connection FINs rate sortedContinuous	0.9503187690510254
Connection PSH rate sortedContinuous	0.7740093815700353
Connection Establishment errors rate sortedContinuous	0.7626191795270896
Connection Other errors rate sortedContinuous	1
Continued on next page	

Table L.52 – continued from previous page

Metric	Similarity
Connection Disconnection errors rate sortedContinuous	0.9962686116548223
Ave duration over last m connections sortedContinuous	0.8851515903313722
Number of packets orderedContinuous	0.274930601454233
Number of packets in orderedContinuous	0.255914693554323
Number of packets out orderedContinuous	0.276832231952716
Duration sortedContinuous	0.9668012824037337
Number control packets rate sortedContinuous	0.9710960544543045
Number data packets rate sortedContinuous	0.959069022124373
Number bytes transferred orderedContinuous	0.214989475919194
Number bytes transferred in orderedContinuous	0.194060148358849
Number bytes transferred out orderedContinuous	0.307035073123228
Number data bytes transferred orderedContinuous	0.204512861791865
Number data bytes transferred in orderedContinuous	0.189960372451521
Number data bytes transferred out orderedContinuous	0.309462124180242
Fragmented packets rate sortedContinuous	0.9827994694098275
Bad fragment rate sortedContinuous	0.9827994694098275
Max Src Window orderedContinuous	0.127940535139985
Max Dst Window orderedContinuous	0.261032858752426
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9687555028345105
Wrong resend rate sortedContinuous	0.9775640201720959
Duplicate ACK rate sortedContinuous	0.9989588139633614
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9997562877038338
Window exceeded rate sortedContinuous	0.9977320643447349
Hole rate sortedContinuous	0.9950591215896519
Number connection errors orderedContinuous	0.469397365168538
Number reset connection orderedContinuous	0.849722582343225
Number other errors orderedContinuous	0.853188496749296
Number disconnection errors orderedContinuous	0.469653335618128
Packet Destination IP nonKeyedSortedContinuous	0.603093354683215
Continued on next page	

Table L.52 – continued from previous page

Metric	Similarity
Bytes Destination IP nonKeyedSortedContinuous	0.516387525459269
Connection Destination IP nonKeyedSortedContinuous	0.830343791347596
Packet Source IP nonKeyedSortedContinuous	0.749757595549097
Bytes Source IP nonKeyedSortedContinuous	0.585830943861306
Connection Source IP nonKeyedSortedContinuous	0.93175366430446

Table L.53: Similarity values for individual metrics of basecase 10,
train pair 1

Metric	Similarity
Packets in count	0.383431239715142
Packets out count	0.0725034072320228
Connections in count	0.433443137990458
Connections out count	0.0266450387396285
Bytes in count	0.337835711264973
Bytes out count	0.139164462776464
SYN-ONLY rate ratio	0.270778118652503
SYN-ACK rate ratio	0.202046536405627
Idle connection rate ratio	0.616292386246711
Half-open connection rate ratio	0.322027399299707
Packet Service discrete	0.080090363394785
Bytes Service discrete	0.0704287095141159
Connection Service discrete	0.0244637608238538
Packet Source port discrete	0.135923535602436
Bytes Source port discrete	0.145726929011325
Connection Source port discrete	0.0965710432921276
Connection Source port orderedContinuous	0.0993977654335274
Packet TTL discrete	0.250344163449898
Packet TTL orderedContinuous	0.328136827230066
Continued on next page	

Table L.53 – continued from previous page

Metric	Similarity
InterPacket delta sortedContinuous	0.4172278450781957
Packet sec orderedContinuous	0.978523717229929
Packet min orderedContinuous	0.891504700058929
Packet GmHour orderedContinuous	0.635835930030981
Packet LocHour orderedContinuous	0.61543065141964
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.955654098121056
Bytes min orderedContinuous	0.856711130145459
Bytes GmHour orderedContinuous	0.7006680179872
Bytes LocHour orderedContinuous	0.657017706398596
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.357714386632655
Packets in last w secs orderedContinuous	0.133996515053501
Priv packets time rate sortedContinuous	0.6527890210993515
Unpriv packets time rate sortedContinuous	0.2589658386076255
Connections time rate sortedContinuous	0.1677008662909785
Priv connections connection time rate sortedContinuous	0.2198824977013851
Unpriv connections connection time rate sortedContinuous	0.02409883053277657
Priv packets priv connection time rate sortedContinuous	0.6984335120678444
Unpriv packets unpriv connection time rate sortedContinuous	0.7627157766043418
SYNs connection time rate sortedContinuous	0.3374649414295446
RSTs connection time rate sortedContinuous	0.1111065492262273
FINs connection time rate sortedContinuous	0.08691427046430119
PSH connection time rate sortedContinuous	0.0244560110285178
Establishment errors connection time rate sortedContinuous	0.3249891741531773
Other errors connection time rate sortedContinuous	0.9949836291896627
Disconnection errors connection time rate sortedContinuous	0.9393755584609443
Ave duration over last w secs sortedContinuous	0.005133495086588748
Priv packets packet rate sortedContinuous	0.7375549655325215
Unpriv packets packet rate sortedContinuous	0.219688862952692
InterConnection delta sortedContinuous	0.2355999351040462

Continued on next page

Table L.53 – continued from previous page

Metric	Similarity
Connection sec orderedContinuous	0.973532596938076
Connection min orderedContinuous	0.774348953053942
Connection GmHour orderedContinuous	0.455098220590541
Connection LocHour orderedContinuous	0.429185657523297
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.7468210854665223
Connection Priv connections rate sortedContinuous	0.5823618129053323
Connection Unpriv connections rate sortedContinuous	0.08273978979547426
Connection Priv packet rate sortedContinuous	0.6483335275347414
Connection Unpriv packet rate sortedContinuous	0.2166272608630156
Connection SYNs rate sortedContinuous	0.7034995214625031
Connection RSTs rate sortedContinuous	0.5803631395519417
Connection FINs rate sortedContinuous	0.6386886138312851
Connection PSH rate sortedContinuous	0.5906666936400806
Connection Establishment errors rate sortedContinuous	0.9231613299270095
Connection Other errors rate sortedContinuous	0.9992544666220321
Connection Disconnection errors rate sortedContinuous	0.99758393817136
Ave duration over last m connections sortedContinuous	0.3686763300345458
Number of packets orderedContinuous	0.0453444642609916
Number of packets in orderedContinuous	0.0438453086747902
Number of packets out orderedContinuous	0.0450433545691789
Duration sortedContinuous	0.8783573539118244
Number control packets rate sortedContinuous	0.9093625446581844
Number data packets rate sortedContinuous	0.9274607882513711
Number bytes transferred orderedContinuous	0.0448544147338474
Number bytes transferred in orderedContinuous	0.0398176722208343
Number bytes transferred out orderedContinuous	0.0661684116486873
Number data bytes transferred orderedContinuous	0.040811251035488
Number data bytes transferred in orderedContinuous	0.0376099592455263
Number data bytes transferred out orderedContinuous	0.0671720369155353
Fragmented packets rate sortedContinuous	0.9990296073006566

Continued on next page

Table L.53 – continued from previous page

Metric	Similarity
Bad fragment rate sortedContinuous	0.9990296073006566
Max Src Window orderedContinuous	0.0921355782013498
Max Dst Window orderedContinuous	0.153627022091627
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9825365791646628
Wrong resend rate sortedContinuous	0.9953792891288993
Duplicate ACK rate sortedContinuous	0.9996613190742972
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9810851570752075
Window exceeded rate sortedContinuous	0.9980184047423792
Hole rate sortedContinuous	0.9985966496237013
Number connection errors orderedContinuous	0.0829509017076571
Number reset connection orderedContinuous	0.506838160959671
Number other errors orderedContinuous	0.316799985223585
Number disconnection errors orderedContinuous	0.471781990717047
Packet Destination IP nonKeyedSortedContinuous	0.827226516166442
Bytes Destination IP nonKeyedSortedContinuous	0.811302954260593
Connection Destination IP nonKeyedSortedContinuous	0.937320302147588
Packet Source IP nonKeyedSortedContinuous	0.585525461047095
Bytes Source IP nonKeyedSortedContinuous	0.335655485885033
Connection Source IP nonKeyedSortedContinuous	0.563365863586695

Table L.54: Similarity values for individual metrics of basecase 10,
test pair 1

Metric	Similarity
Packets in count	0.418590344151991
Packets out count	0.423391082881833
Connections in count	0.792515111244481
Continued on next page	

Table L.54 – continued from previous page

Metric	Similarity
Connections out count	0.533299175112909
Bytes in count	0.337918365099243
Bytes out count	0.500765240453705
SYN-ONLY rate ratio	0.737261756511565
SYN-ACK rate ratio	0.31877443145403
Idle connection rate ratio	0.284297251694024
Half-open connection rate ratio	0.951466012436155
Packet Service discrete	0.127670222358306
Bytes Service discrete	0.0993535105724933
Connection Service discrete	0.0838199632982667
Packet Source port discrete	0.097446733450477
Bytes Source port discrete	0.0958286008979505
Connection Source port discrete	0.135769654231909
Connection Source port orderedContinuous	0.148997183230845
Packet TTL discrete	0.334657209269666
Packet TTL orderedContinuous	0.371125879592716
InterPacket delta sortedContinuous	0.5120857950222139
Packet sec orderedContinuous	0.980251960092796
Packet min orderedContinuous	0.912912897139098
Packet GmHour orderedContinuous	0.736226268450633
Packet LocHour orderedContinuous	0.707090340293925
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.963870393068087
Bytes min orderedContinuous	0.89959945271603
Bytes GmHour orderedContinuous	0.711215702031803
Bytes LocHour orderedContinuous	0.673887481469251
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.5928761590981
Packets in last w secs orderedContinuous	0.356242270018174
Priv packets time rate sortedContinuous	0.6266887540460833
Unpriv packets time rate sortedContinuous	0.867415349000793
Continued on next page	

Table L.54 – continued from previous page

Metric	Similarity
Connections time rate sortedContinuous	0.8389117219139406
Priv connections connection time rate sortedContinuous	0.9154334676091201
Unpriv connections connection time rate sortedContinuous	0.5587848564122655
Priv packets priv connection time rate sortedContinuous	0.5234611112082122
Unpriv packets unpriv connection time rate sortedContinuous	0.7358988644539291
SYNs connection time rate sortedContinuous	0.7912922202393798
RSTs connection time rate sortedContinuous	0.5358487637468163
FINs connection time rate sortedContinuous	0.5773271561077018
PSH connection time rate sortedContinuous	0.7913623748405499
Establishment errors connection time rate sortedContinuous	0.3941114891194356
Other errors connection time rate sortedContinuous	0.9963007985177317
Disconnection errors connection time rate sortedContinuous	0.9763812017891554
Ave duration over last w secs sortedContinuous	0.1672270143635379
Priv packets packet rate sortedContinuous	0.6365532974240112
Unpriv packets packet rate sortedContinuous	0.866320212248582
InterConnection delta sortedContinuous	0.67313964720425
Connection sec orderedContinuous	0.971068420507732
Connection min orderedContinuous	0.718432815611816
Connection GmHour orderedContinuous	0.446125761505441
Connection LocHour orderedContinuous	0.472462581384914
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.7866805126430316
Connection Priv connections rate sortedContinuous	0.7080651587873317
Connection Unpriv connections rate sortedContinuous	0.5649848536501625
Connection Priv packet rate sortedContinuous	0.4350548059248205
Connection Unpriv packet rate sortedContinuous	0.6463136961242468
Connection SYNs rate sortedContinuous	0.7283080253110135
Connection RSTs rate sortedContinuous	0.7928742829497953
Connection FINs rate sortedContinuous	0.7675504725956963
Connection PSH rate sortedContinuous	0.3730607035972858
Connection Establishment errors rate sortedContinuous	0.7304872797097177
Continued on next page	

Table L.54 – continued from previous page

Metric	Similarity
Connection Other errors rate sortedContinuous	0.9994854357861869
Connection Disconnection errors rate sortedContinuous	0.9988234634041959
Ave duration over last m connections sortedContinuous	0.47405952986485
Number of packets orderedContinuous	0.124465426059506
Number of packets in orderedContinuous	0.131737296934052
Number of packets out orderedContinuous	0.12350722350553
Duration sortedContinuous	0.8960719470739451
Number control packets rate sortedContinuous	0.9540008016979921
Number data packets rate sortedContinuous	0.9068495890102393
Number bytes transferred orderedContinuous	0.146176363894684
Number bytes transferred in orderedContinuous	0.139087043311797
Number bytes transferred out orderedContinuous	0.170502226045692
Number data bytes transferred orderedContinuous	0.142190351154465
Number data bytes transferred in orderedContinuous	0.132795258867203
Number data bytes transferred out orderedContinuous	0.167088699095113
Fragmented packets rate sortedContinuous	0.9998318772526973
Bad fragment rate sortedContinuous	0.9998318772526973
Max Src Window orderedContinuous	0.0853985433088034
Max Dst Window orderedContinuous	0.165766560383718
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9657647650102844
Wrong resend rate sortedContinuous	0.973118523296688
Duplicate ACK rate sortedContinuous	0.9990101696837433
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.98036668296192
Window exceeded rate sortedContinuous	0.9946787758285787
Hole rate sortedContinuous	0.9983517149840826
Number connection errors orderedContinuous	0.263076520912734
Number reset connection orderedContinuous	0.878276747763951
Number other errors orderedContinuous	0.837921735779925
Number disconnection errors orderedContinuous	0.646910364778795
Continued on next page	

Table L.54 – continued from previous page

Metric	Similarity
Packet Destination IP nonKeyedSortedContinuous	0.804442142922979
Bytes Destination IP nonKeyedSortedContinuous	0.803940289272819
Connection Destination IP nonKeyedSortedContinuous	0.87118947088171
Packet Source IP nonKeyedSortedContinuous	0.279038981159615
Bytes Source IP nonKeyedSortedContinuous	0.15447146242297
Connection Source IP nonKeyedSortedContinuous	0.246657311846582

Table L.55: Similarity values for individual metrics of basecase 11,
train pair 1

Metric	Similarity
Packets in count	0.957131230184279
Packets out count	0.951811532309413
Connections in count	0.819610522719508
Connections out count	0.933627853160829
Bytes in count	0.831019437542785
Bytes out count	0.588022308623533
SYN-ONLY rate ratio	0.788143822361051
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0.777346682829231
Packet Service discrete	0.00223233208595663
Bytes Service discrete	0.00190802643886312
Connection Service discrete	0.0016681313727032
Packet Source port discrete	0.00719336968695204
Bytes Source port discrete	0.00400627503133335
Connection Source port discrete	0.00750741775174857
Connection Source port orderedContinuous	0.0130513021576292
Packet TTL discrete	0.17720367456153
Continued on next page	

Table L.55 – continued from previous page

Metric	Similarity
Packet TTL orderedContinuous	0.177575723618477
InterPacket delta sortedContinuous	0.5109247125490962
Packet sec orderedContinuous	0.751736133965525
Packet min orderedContinuous	0.610924556123498
Packet GmHour orderedContinuous	0.598888112643986
Packet LocHour orderedContinuous	0.634579661048798
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.705263748797443
Bytes min orderedContinuous	0.378379073247655
Bytes GmHour orderedContinuous	0.487950504449731
Bytes LocHour orderedContinuous	0.511535074417584
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.437108389379664
Packets in last w secs orderedContinuous	0.241312011073699
Priv packets time rate sortedContinuous	0.8861493101507818
Unpriv packets time rate sortedContinuous	0.9235098833603012
Connections time rate sortedContinuous	0.1677818292651076
Priv connections connection time rate sortedContinuous	0.836964727318598
Unpriv connections connection time rate sortedContinuous	0.5452807982182037
Priv packets priv connection time rate sortedContinuous	0.2120171806868645
Unpriv packets unpriv connection time rate sortedContinuous	0.5334487623587025
SYNs connection time rate sortedContinuous	0.4858674003896808
RSTs connection time rate sortedContinuous	0.6862811741163099
FINs connection time rate sortedContinuous	0.4432118304718782
PSH connection time rate sortedContinuous	0.1550690269159148
Establishment errors connection time rate sortedContinuous	0.7619017890089153
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9829528419068831
Ave duration over last w secs sortedContinuous	0.1211974221829654
Priv packets packet rate sortedContinuous	0.8915289602077072
Unpriv packets packet rate sortedContinuous	0.9334921242611493
Continued on next page	

Table L.55 – continued from previous page

Metric	Similarity
InterConnection delta sortedContinuous	0.2128958411908565
Connection sec orderedContinuous	0.714129923193457
Connection min orderedContinuous	0.84031404184776
Connection GmHour orderedContinuous	0.744433608437005
Connection LocHour orderedContinuous	0.77544244690219
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.5397321870716032
Connection Priv connections rate sortedContinuous	0.7894843656293587
Connection Unpriv connections rate sortedContinuous	0.1792096323312921
Connection Priv packet rate sortedContinuous	0.4253254639840076
Connection Unpriv packet rate sortedContinuous	0.4311308782887389
Connection SYNs rate sortedContinuous	0.3619588078729027
Connection RSTs rate sortedContinuous	0.3161387116968518
Connection FINs rate sortedContinuous	0.2780100763140942
Connection PSH rate sortedContinuous	0.4733411534917429
Connection Establishment errors rate sortedContinuous	0.5196105111626088
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9972932005197055
Ave duration over last m connections sortedContinuous	0.2360694486756007
Number of packets orderedContinuous	0.158120689567146
Number of packets in orderedContinuous	0.137854048398597
Number of packets out orderedContinuous	0.144562774512271
Duration sortedContinuous	0.816571640698249
Number control packets rate sortedContinuous	0.8471580797061552
Number data packets rate sortedContinuous	0.9214448997418074
Number bytes transferred orderedContinuous	0.0737068673939479
Number bytes transferred in orderedContinuous	0.0724027475866306
Number bytes transferred out orderedContinuous	0.0869146082344212
Number data bytes transferred orderedContinuous	0.0430062886705788
Number data bytes transferred in orderedContinuous	0.0398658304814341
Number data bytes transferred out orderedContinuous	0.0300178599664245
Continued on next page	

Table L.55 – continued from previous page

Metric	Similarity
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0537545119854229
Max Dst Window orderedContinuous	0.0454706176588017
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.7358206794856553
Wrong resend rate sortedContinuous	0.9902622439708501
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9977426636568849
Window exceeded rate sortedContinuous	0.9144104712172997
Hole rate sortedContinuous	0.9922440238467327
Number connection errors orderedContinuous	0.136944776893025
Number reset connection orderedContinuous	0.407116551577913
Number other errors orderedContinuous	0.401434565625761
Number disconnection errors orderedContinuous	0.489561789217801
Packet Destination IP nonKeyedSortedContinuous	0.602810823925588
Bytes Destination IP nonKeyedSortedContinuous	0.358518696206096
Connection Destination IP nonKeyedSortedContinuous	0.724313723895207
Packet Source IP nonKeyedSortedContinuous	0.705932493613704
Bytes Source IP nonKeyedSortedContinuous	0.551339080596823
Connection Source IP nonKeyedSortedContinuous	0.697366178845254

Table L.56: Similarity values for individual metrics of basecase 11,
train pair 2

Metric	Similarity
Packets in count	0.487704594210612
Packets out count	0.461895482031116
Continued on next page	

Table L.56 – continued from previous page

Metric	Similarity
Connections in count	0.587031290603647
Connections out count	0.822688235272086
Bytes in count	0.621558114320881
Bytes out count	0.178914596981951
SYN-ONLY rate ratio	0.973105755597969
SYN-ACK rate ratio	0.316503700825353
Idle connection rate ratio	0.496635696436934
Half-open connection rate ratio	0.874121285031157
Packet Service discrete	0.121228168257421
Bytes Service discrete	0.0981536654996312
Connection Service discrete	0.0638675837358465
Packet Source port discrete	0.170653992131463
Bytes Source port discrete	0.14448425742828
Connection Source port discrete	0.197377844840871
Connection Source port orderedContinuous	0.218155615920754
Packet TTL discrete	0.4727917327198
Packet TTL orderedContinuous	0.427078390837327
InterPacket delta sortedContinuous	0.3327032957877908
Packet sec orderedContinuous	0.974527364104007
Packet min orderedContinuous	0.787034574368337
Packet GmHour orderedContinuous	0.468941876105518
Packet LocHour orderedContinuous	0.461847232959587
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.964655013714219
Bytes min orderedContinuous	0.799143701894419
Bytes GmHour orderedContinuous	0.393510129872568
Bytes LocHour orderedContinuous	0.379828647711301
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.501781757205538
Packets in last w secs orderedContinuous	0.283388712764721
Priv packets time rate sortedContinuous	0.5525078408841243
Continued on next page	

Table L.56 – continued from previous page

Metric	Similarity
Unpriv packets time rate sortedContinuous	0.8176782011846632
Connections time rate sortedContinuous	0.4552966952107176
Priv connections connection time rate sortedContinuous	0.940828964232376
Unpriv connections connection time rate sortedContinuous	0.6090985387190254
Priv packets priv connection time rate sortedContinuous	0.2215627584097135
Unpriv packets unpriv connection time rate sortedContinuous	0.3041555384124948
SYNs connection time rate sortedContinuous	0.8088468049385665
RSTs connection time rate sortedContinuous	0.6205259074317475
FINs connection time rate sortedContinuous	0.7731553603300414
PSH connection time rate sortedContinuous	0.7164176577751697
Establishment errors connection time rate sortedContinuous	0.5794420171436588
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9946206007134819
Ave duration over last w secs sortedContinuous	0.1960269939657002
Priv packets packet rate sortedContinuous	0.5514313211032532
Unpriv packets packet rate sortedContinuous	0.8266576834278635
InterConnection delta sortedContinuous	0.3814045431863785
Connection sec orderedContinuous	0.978435971055566
Connection min orderedContinuous	0.766592247778816
Connection GmHour orderedContinuous	0.714746057415705
Connection LocHour orderedContinuous	0.691377439624858
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.4547989249863104
Connection Priv connections rate sortedContinuous	0.4200285306391874
Connection Unpriv connections rate sortedContinuous	0.4077745175609512
Connection Priv packet rate sortedContinuous	0.321848090802279
Connection Unpriv packet rate sortedContinuous	0.4023614575584026
Connection SYNs rate sortedContinuous	0.5111226304192661
Connection RSTs rate sortedContinuous	0.5818343019127275
Connection FINs rate sortedContinuous	0.4540763508147313
Connection PSH rate sortedContinuous	0.28878503422385
Continued on next page	

Table L.56 – continued from previous page

Metric	Similarity
Connection Establishment errors rate sortedContinuous	0.6099276747120034
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9951159838454086
Ave duration over last m connections sortedContinuous	0.3742851442312812
Number of packets orderedContinuous	0.159786226864727
Number of packets in orderedContinuous	0.175654982085138
Number of packets out orderedContinuous	0.150677597547075
Duration sortedContinuous	0.8397862183730473
Number control packets rate sortedContinuous	0.9584237109957682
Number data packets rate sortedContinuous	0.9511057380040524
Number bytes transferred orderedContinuous	0.156025435041923
Number bytes transferred in orderedContinuous	0.146408102813396
Number bytes transferred out orderedContinuous	0.222860691929475
Number data bytes transferred orderedContinuous	0.151193574667091
Number data bytes transferred in orderedContinuous	0.142735889953309
Number data bytes transferred out orderedContinuous	0.225712430224233
Fragmented packets rate sortedContinuous	0.9999339014237991
Bad fragment rate sortedContinuous	0.9999339014237991
Max Src Window orderedContinuous	0.126876081750563
Max Dst Window orderedContinuous	0.193487590839114
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9818044947734756
Wrong resend rate sortedContinuous	0.9847100335009843
Duplicate ACK rate sortedContinuous	0.9985901955951354
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9339005344979628
Window exceeded rate sortedContinuous	0.9909237756853962
Hole rate sortedContinuous	0.9992553134955935
Number connection errors orderedContinuous	0.261295167522401
Number reset connection orderedContinuous	0.905763007055263
Number other errors orderedContinuous	0.892704164525554
Continued on next page	

Table L.56 – continued from previous page

Metric	Similarity
Number disconnection errors orderedContinuous	0.73155225433537
Packet Destination IP nonKeyedSortedContinuous	0.882805359118637
Bytes Destination IP nonKeyedSortedContinuous	0.874228475829201
Connection Destination IP nonKeyedSortedContinuous	0.926946914134717
Packet Source IP nonKeyedSortedContinuous	0.733407480981718
Bytes Source IP nonKeyedSortedContinuous	0.677132645479904
Connection Source IP nonKeyedSortedContinuous	0.680358646716754

Table L.57: Similarity values for individual metrics of basecase 11,
train pair 3

Metric	Similarity
Packets in count	0.000195614565949631
Packets out count	0.000164847845438687
Connections in count	0.000356943911131835
Connections out count	0.000165415447257899
Bytes in count	3.36842481979405e-05
Bytes out count	7.91365148662404e-05
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	2.7866826962874e-05
Bytes Service discrete	1.27047137257449e-05
Connection Service discrete	4.41487491044617e-05
Packet Source port discrete	1.72391279920896e-05
Bytes Source port discrete	1.69307872655978e-05
Connection Source port discrete	3.51140663054095e-06
Connection Source port orderedContinuous	2.79127291208429e-05
Continued on next page	

Table L.57 – continued from previous page

Metric	Similarity
Packet TTL discrete	0.00333878797021789
Packet TTL orderedContinuous	0.0157068740566735
InterPacket delta sortedContinuous	0.186031772452163
Packet sec orderedContinuous	0.667572971887114
Packet min orderedContinuous	0.580102387042396
Packet GmHour orderedContinuous	0.505995800869524
Packet LocHour orderedContinuous	0.510925237952591
Packet weekday orderedContinuous	0.4375
Bytes sec orderedContinuous	0.666090117073534
Bytes min orderedContinuous	0.58550538582471
Bytes GmHour orderedContinuous	0.546062493939201
Bytes LocHour orderedContinuous	0.544076958126926
Bytes weekday orderedContinuous	0.4375
Packet size orderedContinuous	0.00532654575927905
Packets in last w secs orderedContinuous	8.68924073420157e-05
Priv packets time rate sortedContinuous	0.4875209262106728
Unpriv packets time rate sortedContinuous	0.6729264776944857
Connections time rate sortedContinuous	0.1720484506272452
Priv connections connection time rate sortedContinuous	0.5436936846716888
Unpriv connections connection time rate sortedContinuous	0.05940816493448073
Priv packets priv connection time rate sortedContinuous	0.283244339614132
Unpriv packets unpriv connection time rate sortedContinuous	0.03594045226892072
SYNs connection time rate sortedContinuous	0.7475731434307608
RSTs connection time rate sortedContinuous	0.3610415168259698
FINs connection time rate sortedContinuous	0.1759952821629468
PSH connection time rate sortedContinuous	0.3346253933373136
Establishment errors connection time rate sortedContinuous	0.4366062940037054
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9857142857142858
Ave duration over last w secs sortedContinuous	0.008414352951155343
Priv packets packet rate sortedContinuous	0.5340075705428185
Continued on next page	

Table L.57 – continued from previous page

Metric	Similarity
Unpriv packets packet rate sortedContinuous	0.8094357746375488
InterConnection delta sortedContinuous	0.1390204037064473
Connection sec orderedContinuous	0.634206400744145
Connection min orderedContinuous	0.560872335444845
Connection GmHour orderedContinuous	0.399342797572342
Connection LocHour orderedContinuous	0.392074664917973
Connection weekday orderedContinuous	0.4375
Connection packet rate sortedContinuous	0.3517990452249006
Connection Priv connections rate sortedContinuous	0.3355175255075084
Connection Unpriv connections rate sortedContinuous	0.3244653007034251
Connection Priv packet rate sortedContinuous	0.3268166888699301
Connection Unpriv packet rate sortedContinuous	0.4898719564534173
Connection SYNs rate sortedContinuous	0.4117460816632646
Connection RSTs rate sortedContinuous	0.3024398929964308
Connection FINs rate sortedContinuous	0.6093511362817009
Connection PSH rate sortedContinuous	0.1500061020784761
Connection Establishment errors rate sortedContinuous	0.3471174717983444
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9782608695652174
Ave duration over last m connections sortedContinuous	0.09867940752435415
Number of packets orderedContinuous	0.00665353210122476
Number of packets in orderedContinuous	0.00744477159255305
Number of packets out orderedContinuous	0.00803139008826209
Duration sortedContinuous	0.7003466948031822
Number control packets rate sortedContinuous	0.8230026117876585
Number data packets rate sortedContinuous	0.8425451139604702
Number bytes transferred orderedContinuous	4.88869008623292e-05
Number bytes transferred in orderedContinuous	0.000276742588456895
Number bytes transferred out orderedContinuous	0.00083459912254684
Number data bytes transferred orderedContinuous	0.000172473255140916
Number data bytes transferred in orderedContinuous	0.000209381180772931
Continued on next page	

Table L.57 – continued from previous page

Metric	Similarity
Number data bytes transferred out orderedContinuous	0.000511842690552161
Fragmented packets rate sortedContinuous	0.9782608695652174
Bad fragment rate sortedContinuous	0.9782608695652174
Max Src Window orderedContinuous	0.00068664145709771
Max Dst Window orderedContinuous	0.00150810812493236
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.8869701726844585
Wrong resend rate sortedContinuous	0.941872941223659
Duplicate ACK rate sortedContinuous	0.9285714285714286
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9285714285714286
Window exceeded rate sortedContinuous	0.9285714285714286
Hole rate sortedContinuous	0.9285714285714286
Number connection errors orderedContinuous	0.0913352797348639
Number reset connection orderedContinuous	0.640469122713871
Number other errors orderedContinuous	0.649497663528339
Number disconnection errors orderedContinuous	0.3683116780951
Packet Destination IP nonKeyedSortedContinuous	0.456985104182802
Bytes Destination IP nonKeyedSortedContinuous	0.399440618787958
Connection Destination IP nonKeyedSortedContinuous	0.845290984515195
Packet Source IP nonKeyedSortedContinuous	0.165052263201572
Bytes Source IP nonKeyedSortedContinuous	0.162333783055111
Connection Source IP nonKeyedSortedContinuous	0.220446123502836

Table L.58: Similarity values for individual metrics of basecase 11,
test pair 1

Metric	Similarity
Packets in count	0.0418167797430664
Continued on next page	

Table L.58 – continued from previous page

Metric	Similarity
Packets out count	0.037138384964472
Connections in count	0.0154527198209585
Connections out count	0.00066624269404314
Bytes in count	0.0382627155035349
Bytes out count	0.014284414804176
SYN-ONLY rate ratio	0.0775153531399265
SYN-ACK rate ratio	0.924750424588874
Idle connection rate ratio	0.337069389336101
Half-open connection rate ratio	0.935485909720992
Packet Service discrete	0.0103348761461262
Bytes Service discrete	0.0152513902919844
Connection Service discrete	0.000373912889591867
Packet Source port discrete	0.0105691613810008
Bytes Source port discrete	0.0145389051906893
Connection Source port discrete	0.00109173710411503
Connection Source port orderedContinuous	0.00335792668192765
Packet TTL discrete	0.0623866962932865
Packet TTL orderedContinuous	0.12776167182405
InterPacket delta sortedContinuous	0.3758461582084971
Packet sec orderedContinuous	0.915612533631816
Packet min orderedContinuous	0.531509626966193
Packet GmHour orderedContinuous	0.110712590010764
Packet LocHour orderedContinuous	0.0896430551650794
Packet weekday orderedContinuous	0.464568484231988
Bytes sec orderedContinuous	0.871055132574723
Bytes min orderedContinuous	0.273963322211467
Bytes GmHour orderedContinuous	0.126313547431106
Bytes LocHour orderedContinuous	0.112184039374487
Bytes weekday orderedContinuous	0.462629716236163
Packet size orderedContinuous	0.611806668962194
Packets in last w secs orderedContinuous	0.286779592151298

Continued on next page

Table L.58 – continued from previous page

Metric	Similarity
Priv packets time rate sortedContinuous	0.7065787643156577
Unpriv packets time rate sortedContinuous	0.9270482896290573
Connections time rate sortedContinuous	0.2773833864500812
Priv connections connection time rate sortedContinuous	0.5758579606719531
Unpriv connections connection time rate sortedContinuous	0.06336876782892065
Priv packets priv connection time rate sortedContinuous	0.2425226099812658
Unpriv packets unpriv connection time rate sortedContinuous	0.03653337192620322
SYNs connection time rate sortedContinuous	0.7248894857082611
RSTs connection time rate sortedContinuous	0.4355213982570535
FINs connection time rate sortedContinuous	0.4388744284140562
PSH connection time rate sortedContinuous	0.1909560230920055
Establishment errors connection time rate sortedContinuous	0.3743044138346716
Other errors connection time rate sortedContinuous	0.9958789922262808
Disconnection errors connection time rate sortedContinuous	0.8859854515937685
Ave duration over last w secs sortedContinuous	0.2917521902395133
Priv packets packet rate sortedContinuous	0.7283855906392018
Unpriv packets packet rate sortedContinuous	0.9232459275647837
InterConnection delta sortedContinuous	0.2790559562078559
Connection sec orderedContinuous	0.761959049091772
Connection min orderedContinuous	0.48577994114476
Connection GmHour orderedContinuous	0.091833875596892
Connection LocHour orderedContinuous	0.106678983557953
Connection weekday orderedContinuous	0.464041995549804
Connection packet rate sortedContinuous	0.3190617612196914
Connection Priv connections rate sortedContinuous	0.2624770156842376
Connection Unpriv connections rate sortedContinuous	0.1761998001996323
Connection Priv packet rate sortedContinuous	0.1486932504865394
Connection Unpriv packet rate sortedContinuous	0.2712810821330174
Connection SYNs rate sortedContinuous	0.2663529670186547
Connection RSTs rate sortedContinuous	0.6289825184489258
Connection FINs rate sortedContinuous	0.2494642430027938
Continued on next page	

Table L.58 – continued from previous page

Metric	Similarity
Connection PSH rate sortedContinuous	0.2161064534502363
Connection Establishment errors rate sortedContinuous	0.5500804316157925
Connection Other errors rate sortedContinuous	0.9939556749496307
Connection Disconnection errors rate sortedContinuous	0.9905977165883143
Ave duration over last m connections sortedContinuous	0.4487422534383609
Number of packets orderedContinuous	0.0261718768048155
Number of packets in orderedContinuous	0.0259686788645066
Number of packets out orderedContinuous	0.0240312817134167
Duration sortedContinuous	0.05471930175142683
Number control packets rate sortedContinuous	0.1171469878109609
Number data packets rate sortedContinuous	0.2355639847136187
Number bytes transferred orderedContinuous	0.00747181169779753
Number bytes transferred in orderedContinuous	0.00602478496310333
Number bytes transferred out orderedContinuous	0.00947775571012775
Number data bytes transferred orderedContinuous	0.000784599646283183
Number data bytes transferred in orderedContinuous	0.00102743395186073
Number data bytes transferred out orderedContinuous	0.00524895773114564
Fragmented packets rate sortedContinuous	0.9818670248488919
Bad fragment rate sortedContinuous	0.9818670248488919
Max Src Window orderedContinuous	0.0142610703008469
Max Dst Window orderedContinuous	0.00452952429981189
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.6440573669103813
Wrong resend rate sortedContinuous	0.8926932812869333
Duplicate ACK rate sortedContinuous	0.9971830985915493
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9988262910798122
Window exceeded rate sortedContinuous	0.9957571642296169
Hole rate sortedContinuous	0.9574773868429946
Number connection errors orderedContinuous	0.115871096995409
Number reset connection orderedContinuous	0.156966988262092
Continued on next page	

Table L.58 – continued from previous page

Metric	Similarity
Number other errors orderedContinuous	0.181048470064304
Number disconnection errors orderedContinuous	0.0571173237746175
Packet Destination IP nonKeyedSortedContinuous	0.712618249895111
Bytes Destination IP nonKeyedSortedContinuous	0.569760460004163
Connection Destination IP nonKeyedSortedContinuous	0.823320526686862
Packet Source IP nonKeyedSortedContinuous	0.814419441078411
Bytes Source IP nonKeyedSortedContinuous	0.73291314765043
Connection Source IP nonKeyedSortedContinuous	0.52885511255716

Table L.59: Similarity values for individual metrics of basecase 11,
test pair 2

Metric	Similarity
Packets in count	0.647883016185296
Packets out count	0.620618035071827
Connections in count	0.920826352688448
Connections out count	0.86990958678921
Bytes in count	0.630816180894325
Bytes out count	0.347098177773066
SYN-ONLY rate ratio	0.660406722720659
SYN-ACK rate ratio	0.345733421172739
Idle connection rate ratio	0.4172897967571
Half-open connection rate ratio	0.671097460833717
Packet Service discrete	0.110412224499812
Bytes Service discrete	0.0880388125219761
Connection Service discrete	0.0628405617200099
Packet Source port discrete	0.0910871331292897
Bytes Source port discrete	0.0725406155267116
Connection Source port discrete	0.0972774022673171
Continued on next page	

Table L.59 – continued from previous page

Metric	Similarity
Connection Source port orderedContinuous	0.128196281027332
Packet TTL discrete	0.200949362457288
Packet TTL orderedContinuous	0.285168356189357
InterPacket delta sortedContinuous	0.7806055504244304
Packet sec orderedContinuous	0.974480421574872
Packet min orderedContinuous	0.845550924736109
Packet GmHour orderedContinuous	0.687428205010892
Packet LocHour orderedContinuous	0.677558232663493
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.955213459089851
Bytes min orderedContinuous	0.713712137897059
Bytes GmHour orderedContinuous	0.607954896095792
Bytes LocHour orderedContinuous	0.597167895637372
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.581583032298895
Packets in last w secs orderedContinuous	0.50868939476589
Priv packets time rate sortedContinuous	0.8307983122504709
Unpriv packets time rate sortedContinuous	0.9239473882353754
Connections time rate sortedContinuous	0.7298806052509364
Priv connections connection time rate sortedContinuous	0.9744547488868571
Unpriv connections connection time rate sortedContinuous	0.9214724418235965
Priv packets priv connection time rate sortedContinuous	0.7055984794966832
Unpriv packets unpriv connection time rate sortedContinuous	0.7258010173243749
SYNs connection time rate sortedContinuous	0.7346920900018092
RSTs connection time rate sortedContinuous	0.7357454753013195
FINs connection time rate sortedContinuous	0.8836722004477958
PSH connection time rate sortedContinuous	0.3107958225790095
Establishment errors connection time rate sortedContinuous	0.8600765705750882
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.991878287096158
Ave duration over last w secs sortedContinuous	0.02546189175789409
Continued on next page	

Table L.59 – continued from previous page

Metric	Similarity
Priv packets packet rate sortedContinuous	0.8722741447940871
Unpriv packets packet rate sortedContinuous	0.9256191789919084
InterConnection delta sortedContinuous	0.5093314643564398
Connection sec orderedContinuous	0.96775947229546
Connection min orderedContinuous	0.745293667181295
Connection GmHour orderedContinuous	0.63072098106458
Connection LocHour orderedContinuous	0.634031988577446
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.4591456801744281
Connection Priv connections rate sortedContinuous	0.8270670498305798
Connection Unpriv connections rate sortedContinuous	0.7396889674500346
Connection Priv packet rate sortedContinuous	0.6275863038587343
Connection Unpriv packet rate sortedContinuous	0.4693896922179941
Connection SYNs rate sortedContinuous	0.5629317618806158
Connection RSTs rate sortedContinuous	0.6563086426380266
Connection FINs rate sortedContinuous	0.8344431908842168
Connection PSH rate sortedContinuous	0.4409737375803208
Connection Establishment errors rate sortedContinuous	0.7775704561300977
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9985523910694931
Ave duration over last m connections sortedContinuous	0.3876417794347048
Number of packets orderedContinuous	0.172509373106564
Number of packets in orderedContinuous	0.154918448349088
Number of packets out orderedContinuous	0.128735357207206
Duration sortedContinuous	0.8781944072372385
Number control packets rate sortedContinuous	0.8892152514575114
Number data packets rate sortedContinuous	0.8851380987593378
Number bytes transferred orderedContinuous	0.157228128715571
Number bytes transferred in orderedContinuous	0.151756744432364
Number bytes transferred out orderedContinuous	0.251370772545239
Number data bytes transferred orderedContinuous	0.155958224169034
Continued on next page	

Table L.59 – continued from previous page

Metric	Similarity
Number data bytes transferred in orderedContinuous	0.142870843409105
Number data bytes transferred out orderedContinuous	0.248083533212497
Fragmented packets rate sortedContinuous	0.9998643219934049
Bad fragment rate sortedContinuous	0.9998643219934049
Max Src Window orderedContinuous	0.132637353365107
Max Dst Window orderedContinuous	0.153178282326162
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9556890260762244
Wrong resend rate sortedContinuous	0.9591589317585418
Duplicate ACK rate sortedContinuous	0.9993863869300817
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.6772143340538473
Window exceeded rate sortedContinuous	0.997552297800074
Hole rate sortedContinuous	0.9945054339704838
Number connection errors orderedContinuous	0.213234829002017
Number reset connection orderedContinuous	0.46445334544514
Number other errors orderedContinuous	0.480269775834671
Number disconnection errors orderedContinuous	0.29594787268308
Packet Destination IP nonKeyedSortedContinuous	0.360200002971698
Bytes Destination IP nonKeyedSortedContinuous	0.347230558576136
Connection Destination IP nonKeyedSortedContinuous	0.589972743423545
Packet Source IP nonKeyedSortedContinuous	0.559342479601075
Bytes Source IP nonKeyedSortedContinuous	0.428434865075266
Connection Source IP nonKeyedSortedContinuous	0.506633937937546

Table L.60: Similarity values for individual metrics of basecase 11,
test pair 3

Metric	Similarity
Packets in count	0.824731452153685
Packets out count	0.863553568745056
Connections in count	0.288479688915735
Connections out count	0.794073440076822
Bytes in count	0.947009895888172
Bytes out count	0.932309650051063
SYN-ONLY rate ratio	0.0635876093349974
SYN-ACK rate ratio	0.00590797005855881
Idle connection rate ratio	0.0678641985195678
Half-open connection rate ratio	0.107243892580069
Packet Service discrete	0.110240488597697
Bytes Service discrete	0.0884658656094113
Connection Service discrete	0.0609479664847255
Packet Source port discrete	0.136239747383339
Bytes Source port discrete	0.136555391588189
Connection Source port discrete	0.113151402445467
Connection Source port orderedContinuous	0.136554591978747
Packet TTL discrete	0.236183769699161
Packet TTL orderedContinuous	0.312392052143855
InterPacket delta sortedContinuous	0.3570066798890081
Packet sec orderedContinuous	0.988670282698888
Packet min orderedContinuous	0.886286626676815
Packet GmHour orderedContinuous	0.457666353727741
Packet LocHour orderedContinuous	0.365644671412522
Packet weekday orderedContinuous	0.81519985711678
Bytes sec orderedContinuous	0.971958503063681
Bytes min orderedContinuous	0.758897384160561
Bytes GmHour orderedContinuous	0.562987741440282
Bytes LocHour orderedContinuous	0.437933107851159
Continued on next page	

Table L.60 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	0.835830965737151
Packet size orderedContinuous	0.36965871607093
Packets in last w secs orderedContinuous	0.22802604746513
Priv packets time rate sortedContinuous	0.4933244833879888
Unpriv packets time rate sortedContinuous	0.8079885786708496
Connections time rate sortedContinuous	0.7476085796758083
Priv connections connection time rate sortedContinuous	0.5916213914947158
Unpriv connections connection time rate sortedContinuous	0.4517614641393447
Priv packets priv connection time rate sortedContinuous	0.6129036054336933
Unpriv packets unpriv connection time rate sortedContinuous	0.5110049803673743
SYNs connection time rate sortedContinuous	0.6706199901484916
RSTs connection time rate sortedContinuous	0.6259971756978247
FINs connection time rate sortedContinuous	0.5605128013837066
PSH connection time rate sortedContinuous	0.6680273978331565
Establishment errors connection time rate sortedContinuous	0.3595794058998857
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.8911187121591145
Ave duration over last w secs sortedContinuous	0.4939677456186399
Priv packets packet rate sortedContinuous	0.4410355009320061
Unpriv packets packet rate sortedContinuous	0.7992220879570655
InterConnection delta sortedContinuous	0.610331854149796
Connection sec orderedContinuous	0.956722342848234
Connection min orderedContinuous	0.800462177799768
Connection GmHour orderedContinuous	0.403638242714047
Connection LocHour orderedContinuous	0.364057521911397
Connection weekday orderedContinuous	0.817385534373289
Connection packet rate sortedContinuous	0.6995696935395166
Connection Priv connections rate sortedContinuous	0.6853413477479049
Connection Unpriv connections rate sortedContinuous	0.4655178022244852
Connection Priv packet rate sortedContinuous	0.4100353081022868
Connection Unpriv packet rate sortedContinuous	0.3974650613330883
Continued on next page	

Table L.60 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.4295350839316975
Connection RSTs rate sortedContinuous	0.5790835391766926
Connection FINs rate sortedContinuous	0.2769550672892007
Connection PSH rate sortedContinuous	0.6216359048245585
Connection Establishment errors rate sortedContinuous	0.7684175318827953
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9945039843744959
Ave duration over last m connections sortedContinuous	0.4644068245289983
Number of packets orderedContinuous	0.231942483280657
Number of packets in orderedContinuous	0.22755413556157
Number of packets out orderedContinuous	0.256963526403906
Duration sortedContinuous	0.3852224390626197
Number control packets rate sortedContinuous	0.4078131955206546
Number data packets rate sortedContinuous	0.5504332857600461
Number bytes transferred orderedContinuous	0.153956966713716
Number bytes transferred in orderedContinuous	0.134578507501638
Number bytes transferred out orderedContinuous	0.245487284404332
Number data bytes transferred orderedContinuous	0.155905219973982
Number data bytes transferred in orderedContinuous	0.141181364523608
Number data bytes transferred out orderedContinuous	0.252374776170404
Fragmented packets rate sortedContinuous	0.9834816590588158
Bad fragment rate sortedContinuous	0.9834816590588158
Max Src Window orderedContinuous	0.0235239752811395
Max Dst Window orderedContinuous	0.150265098747466
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.8586163025406513
Wrong resend rate sortedContinuous	0.8935112622647108
Duplicate ACK rate sortedContinuous	0.9976113288841501
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9989149827168746
Window exceeded rate sortedContinuous	0.9951805005753207
Continued on next page	

Table L.60 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9877244338312601
Number connection errors orderedContinuous	0.0403488939983127
Number reset connection orderedContinuous	0.243069875046289
Number other errors orderedContinuous	0.214247659868192
Number disconnection errors orderedContinuous	0.276907235420551
Packet Destination IP nonKeyedSortedContinuous	0.305152204276787
Bytes Destination IP nonKeyedSortedContinuous	0.224623749815975
Connection Destination IP nonKeyedSortedContinuous	0.649769246022629
Packet Source IP nonKeyedSortedContinuous	0.383360792245957
Bytes Source IP nonKeyedSortedContinuous	0.193983943421532
Connection Source IP nonKeyedSortedContinuous	0.603159538554214

Table L.61: Similarity values for individual metrics of basecase 12,
train pair 1

Metric	Similarity
Packets in count	0.0327799846779789
Packets out count	0.0352523267615367
Connections in count	0.0544095979793727
Connections out count	0.0583962345732393
Bytes in count	0.0102999463622108
Bytes out count	0.0403878406348116
SYN-ONLY rate ratio	0.783507717616324
SYN-ACK rate ratio	0
Idle connection rate ratio	0.215878870491387
Half-open connection rate ratio	0.864792301841678
Packet Service discrete	0.000793804188736888
Bytes Service discrete	0.000957541878225119
Connection Service discrete	0.000228707477515772
Continued on next page	

Table L.61 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.00169297352950866
Bytes Source port discrete	0.0021724883291119
Connection Source port discrete	0.00350736641694339
Connection Source port orderedContinuous	0.0232982004027776
Packet TTL discrete	0.0968310377262274
Packet TTL orderedContinuous	0.180775219219099
InterPacket delta sortedContinuous	0.5373095777641092
Packet sec orderedContinuous	0.907492872722
Packet min orderedContinuous	0.783513537717169
Packet GmHour orderedContinuous	0.659276391136957
Packet LocHour orderedContinuous	0.637932080372685
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.785461093192399
Bytes min orderedContinuous	0.688040235693486
Bytes GmHour orderedContinuous	0.551514759613739
Bytes LocHour orderedContinuous	0.527497333040837
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.422312501772192
Packets in last w secs orderedContinuous	0.0795635934140632
Priv packets time rate sortedContinuous	0.535633499242692
Unpriv packets time rate sortedContinuous	0.7976491973202223
Connections time rate sortedContinuous	0.7362245212728739
Priv connections connection time rate sortedContinuous	0.5550768953990095
Unpriv connections connection time rate sortedContinuous	0.03134340296813362
Priv packets priv connection time rate sortedContinuous	0.8519543811085483
Unpriv packets unpriv connection time rate sortedContinuous	0.02867534043611327
SYNs connection time rate sortedContinuous	0.6082885553798765
RSTs connection time rate sortedContinuous	0.4559132774651723
FINs connection time rate sortedContinuous	0.508950416208448
PSH connection time rate sortedContinuous	0.6151304117355775
Establishment errors connection time rate sortedContinuous	0.396504331154664
Continued on next page	

Table L.61 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.989658273381295
Ave duration over last w secs sortedContinuous	0.002807391803217544
Priv packets packet rate sortedContinuous	0.53890338707898
Unpriv packets packet rate sortedContinuous	0.8222854061302657
InterConnection delta sortedContinuous	0.3981940825086077
Connection sec orderedContinuous	0.913850198623047
Connection min orderedContinuous	0.690227634449178
Connection GmHour orderedContinuous	0.57681953595836
Connection LocHour orderedContinuous	0.548622809231118
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.4484856741845676
Connection Priv connections rate sortedContinuous	0.3227168624336531
Connection Unpriv connections rate sortedContinuous	0.07685852006897116
Connection Priv packet rate sortedContinuous	0.2578573476124383
Connection Unpriv packet rate sortedContinuous	0.4627786583512604
Connection SYNs rate sortedContinuous	0.4628517443978264
Connection RSTs rate sortedContinuous	0.7466157419531475
Connection FINs rate sortedContinuous	0.5044895312615191
Connection PSH rate sortedContinuous	0.2077068311821385
Connection Establishment errors rate sortedContinuous	0.68098448166495
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9972719005759321
Ave duration over last m connections sortedContinuous	0.1283226191687004
Number of packets orderedContinuous	0.0430380433615877
Number of packets in orderedContinuous	0.0373618958011464
Number of packets out orderedContinuous	0.0367580396613831
Duration sortedContinuous	0.8854516661351403
Number control packets rate sortedContinuous	0.898194876193061
Number data packets rate sortedContinuous	0.8877719838753297
Number bytes transferred orderedContinuous	0.0194153298268392
Continued on next page	

Table L.61 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.0193937404296412
Number bytes transferred out orderedContinuous	0.0466703489341869
Number data bytes transferred orderedContinuous	0.0148217749050309
Number data bytes transferred in orderedContinuous	0.0123454206638262
Number data bytes transferred out orderedContinuous	0.0180426657473702
Fragmented packets rate sortedContinuous	0.9998989286436224
Bad fragment rate sortedContinuous	0.9998989286436224
Max Src Window orderedContinuous	0.0129282349186853
Max Dst Window orderedContinuous	0.0249028849361447
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9120869858108026
Wrong resend rate sortedContinuous	0.9710236990429301
Duplicate ACK rate sortedContinuous	0.9983079526226735
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9807106598984771
Window exceeded rate sortedContinuous	0.9315743398303553
Hole rate sortedContinuous	0.9984541789392326
Number connection errors orderedContinuous	0.0775975579772458
Number reset connection orderedContinuous	0.552714528414916
Number other errors orderedContinuous	0.579518244896603
Number disconnection errors orderedContinuous	0.566042291937102
Packet Destination IP nonKeyedSortedContinuous	0.264986467068589
Bytes Destination IP nonKeyedSortedContinuous	0.305311552020059
Connection Destination IP nonKeyedSortedContinuous	0.444447981855167
Packet Source IP nonKeyedSortedContinuous	0.176418259008558
Bytes Source IP nonKeyedSortedContinuous	0.111706057306724
Connection Source IP nonKeyedSortedContinuous	0.193268744379621

Table L.62: Similarity values for individual metrics of basecase 12,
train pair 2

Metric	Similarity
Packets in count	0.00719748088169136
Packets out count	0.00559284116331094
Connections in count	0.00686713141192385
Connections out count	0.00290340591848126
Bytes in count	0.00445639179471258
Bytes out count	0.00342036230372011
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0
Packet Service discrete	3.61455551016707e-05
Bytes Service discrete	2.65241876848927e-05
Connection Service discrete	1.27630046663215e-05
Packet Source port discrete	3.41517130345835e-05
Bytes Source port discrete	5.2093630939629e-05
Connection Source port discrete	3.42512156528948e-06
Connection Source port orderedContinuous	4.13701722136683e-05
Packet TTL discrete	0.0624708902280947
Packet TTL orderedContinuous	0.0556923151107068
InterPacket delta sortedContinuous	0.1092397986105857
Packet sec orderedContinuous	0.648996636591286
Packet min orderedContinuous	0.559539288329413
Packet GmHour orderedContinuous	0.687121008327613
Packet LocHour orderedContinuous	0.677575028223949
Packet weekday orderedContinuous	1
Bytes sec orderedContinuous	0.622717206951015
Bytes min orderedContinuous	0.540536747720262
Bytes GmHour orderedContinuous	0.609142831587997
Bytes LocHour orderedContinuous	0.601892113008157
Continued on next page	

Table L.62 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	1
Packet size orderedContinuous	0.00879780149812149
Packets in last w secs orderedContinuous	0.00262078922962528
Priv packets time rate sortedContinuous	0.8868873841949048
Unpriv packets time rate sortedContinuous	0.8371775489045694
Connections time rate sortedContinuous	0.2633524236307703
Priv connections connection time rate sortedContinuous	0.9472871572871573
Unpriv connections connection time rate sortedContinuous	0.9650680272108844
Priv packets priv connection time rate sortedContinuous	0.3058731949042436
Unpriv packets unpriv connection time rate sortedContinuous	0.9527733864597022
SYNs connection time rate sortedContinuous	0.6925078715908415
RSTs connection time rate sortedContinuous	0.8014976452119308
FINs connection time rate sortedContinuous	0.5401660025189436
PSH connection time rate sortedContinuous	0.3717653961396622
Establishment errors connection time rate sortedContinuous	0.9615414347557204
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.1273331621413883
Priv packets packet rate sortedContinuous	0.9720715131267097
Unpriv packets packet rate sortedContinuous	0.9772974282256125
InterConnection delta sortedContinuous	0.06347910704381476
Connection sec orderedContinuous	0.625009074531282
Connection min orderedContinuous	0.610593651397105
Connection GmHour orderedContinuous	0.689866143171457
Connection LocHour orderedContinuous	0.696670237845841
Connection weekday orderedContinuous	1
Connection packet rate sortedContinuous	0.8335714129161711
Connection Priv connections rate sortedContinuous	0.9510138091101849
Connection Unpriv connections rate sortedContinuous	0.3637622272763386
Connection Priv packet rate sortedContinuous	0.801150063781225
Connection Unpriv packet rate sortedContinuous	0.4025166653058436

Continued on next page

Table L.62 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.387641327908905
Connection RSTs rate sortedContinuous	0.2055025917011243
Connection FINs rate sortedContinuous	0.1998327759197324
Connection PSH rate sortedContinuous	0.6810498451926935
Connection Establishment errors rate sortedContinuous	0.2066584888092736
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.4142901753442627
Number of packets orderedContinuous	0.0360930962713676
Number of packets in orderedContinuous	0.0829109255266011
Number of packets out orderedContinuous	0.0743244522555932
Duration sortedContinuous	0.706404818206762
Number control packets rate sortedContinuous	0.7520507787503081
Number data packets rate sortedContinuous	0.8937886114737886
Number bytes transferred orderedContinuous	0.00125588998572347
Number bytes transferred in orderedContinuous	0.00140047881618184
Number bytes transferred out orderedContinuous	0.00213898209203819
Number data bytes transferred orderedContinuous	0.00116712174263077
Number data bytes transferred in orderedContinuous	0.00145920238079024
Number data bytes transferred out orderedContinuous	0.00338893085845241
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0227604565445772
Max Dst Window orderedContinuous	0.000877939853489269
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9231309231309232
Wrong resend rate sortedContinuous	0.973469387755102
Duplicate ACK rate sortedContinuous	0.9285714285714286
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9285714285714286
Window exceeded rate sortedContinuous	0.9285714285714286
Continued on next page	

Table L.62 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9285714285714286
Number connection errors orderedContinuous	0.144177737987623
Number reset connection orderedContinuous	0.258049403815752
Number other errors orderedContinuous	0.266091550473133
Number disconnection errors orderedContinuous	0.266091550473133
Packet Destination IP nonKeyedSortedContinuous	0.469752704006557
Bytes Destination IP nonKeyedSortedContinuous	0.738969828221765
Connection Destination IP nonKeyedSortedContinuous	0.536162536545168
Packet Source IP nonKeyedSortedContinuous	0.511148102598479
Bytes Source IP nonKeyedSortedContinuous	0.426786320838803
Connection Source IP nonKeyedSortedContinuous	0.811820284684606

Table L.63: Similarity values for individual metrics of basecase 12,
train pair 3

Metric	Similarity
Packets in count	0.280559532988215
Packets out count	0.21663923806856
Connections in count	0.294775149467117
Connections out count	0.232844482785664
Bytes in count	0.45824782608941
Bytes out count	0.197378201699236
SYN-ONLY rate ratio	0.625424600620717
SYN-ACK rate ratio	0
Idle connection rate ratio	0.754463250637608
Half-open connection rate ratio	0.770166438878095
Packet Service discrete	0.00285928245508156
Bytes Service discrete	0.00338343276406517
Connection Service discrete	0.00245214133784473
Continued on next page	

Table L.63 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.00332946679871812
Bytes Source port discrete	0.00387471722593335
Connection Source port discrete	0.0137161934762108
Connection Source port orderedContinuous	0.0609162827218471
Packet TTL discrete	0.0441946000384335
Packet TTL orderedContinuous	0.119277684716431
InterPacket delta sortedContinuous	0.3750088495045373
Packet sec orderedContinuous	0.840811072961986
Packet min orderedContinuous	0.701141871383093
Packet GmHour orderedContinuous	0.297382756322286
Packet LocHour orderedContinuous	0.273894029803949
Packet weekday orderedContinuous	0.839222265129051
Bytes sec orderedContinuous	0.839452912039441
Bytes min orderedContinuous	0.67755464379643
Bytes GmHour orderedContinuous	0.296749303605355
Bytes LocHour orderedContinuous	0.263672433857095
Bytes weekday orderedContinuous	0.786895786066883
Packet size orderedContinuous	0.239494599794954
Packets in last w secs orderedContinuous	0.0154863435129477
Priv packets time rate sortedContinuous	0.8758367939614993
Unpriv packets time rate sortedContinuous	0.8945305993412184
Connections time rate sortedContinuous	0.7654952757854133
Priv connections connection time rate sortedContinuous	0.7823624074082531
Unpriv connections connection time rate sortedContinuous	0.5833136664848086
Priv packets priv connection time rate sortedContinuous	0.793179415188525
Unpriv packets unpriv connection time rate sortedContinuous	0.5814242492742581
SYNs connection time rate sortedContinuous	0.6530519911507912
RSTs connection time rate sortedContinuous	0.5565533127083355
FINs connection time rate sortedContinuous	0.6186708187634091
PSH connection time rate sortedContinuous	0.4830066236127302
Establishment errors connection time rate sortedContinuous	0.6002861115849579
Continued on next page	

Table L.63 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.5178183634798243
Priv packets packet rate sortedContinuous	0.8006326754476645
Unpriv packets packet rate sortedContinuous	0.845758483671124
InterConnection delta sortedContinuous	0.3014808185015687
Connection sec orderedContinuous	0.727336047437003
Connection min orderedContinuous	0.574439830518409
Connection GmHour orderedContinuous	0.29663906464718
Connection LocHour orderedContinuous	0.267857237783694
Connection weekday orderedContinuous	0.935127139581752
Connection packet rate sortedContinuous	0.8567070251662339
Connection Priv connections rate sortedContinuous	0.6168249864410384
Connection Unpriv connections rate sortedContinuous	0.4994820617658157
Connection Priv packet rate sortedContinuous	0.619409243338345
Connection Unpriv packet rate sortedContinuous	0.8058616130163996
Connection SYNs rate sortedContinuous	0.301249871843336
Connection RSTs rate sortedContinuous	0.406100319226126
Connection FINs rate sortedContinuous	0.3720012954619448
Connection PSH rate sortedContinuous	0.2763562242401184
Connection Establishment errors rate sortedContinuous	0.4203695462698689
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.26719776295204
Number of packets orderedContinuous	0.167348559034203
Number of packets in orderedContinuous	0.206721952060229
Number of packets out orderedContinuous	0.140261184447878
Duration sortedContinuous	0.377230665403385
Number control packets rate sortedContinuous	0.3326699581374435
Number data packets rate sortedContinuous	0.3481674457717039
Number bytes transferred orderedContinuous	0.0572717624761094
Continued on next page	

Table L.63 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.0629038678153553
Number bytes transferred out orderedContinuous	0.100481743479463
Number data bytes transferred orderedContinuous	0.0394901790019485
Number data bytes transferred in orderedContinuous	0.0456954463872899
Number data bytes transferred out orderedContinuous	0.0475085343501696
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.00353266142437955
Max Dst Window orderedContinuous	0.0180073635012276
Urgent rate sortedContinuous	0.9968354430379747
Resend rate sortedContinuous	0.8649864584749994
Wrong resend rate sortedContinuous	0.9644452717795979
Duplicate ACK rate sortedContinuous	0.9997123130034522
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9968354430379747
Window exceeded rate sortedContinuous	0.9530001863823039
Hole rate sortedContinuous	0.9968354430379747
Number connection errors orderedContinuous	0.188840318637487
Number reset connection orderedContinuous	0.127205043322066
Number other errors orderedContinuous	0.133973467945504
Number disconnection errors orderedContinuous	0.133973467945504
Packet Destination IP nonKeyedSortedContinuous	0.294801840135459
Bytes Destination IP nonKeyedSortedContinuous	0.169713355021656
Connection Destination IP nonKeyedSortedContinuous	0.630883240156182
Packet Source IP nonKeyedSortedContinuous	0.549777011530985
Bytes Source IP nonKeyedSortedContinuous	0.390283197328719
Connection Source IP nonKeyedSortedContinuous	0.27560719462028

Table L.64: Similarity values for individual metrics of basecase 12,
test pair 1

Metric	Similarity
Packets in count	0.196784301079864
Packets out count	0.164425666899107
Connections in count	0.0636460599340761
Connections out count	0.041130353381699
Bytes in count	0.200382399044708
Bytes out count	0.142241019739555
SYN-ONLY rate ratio	0.798475083458737
SYN-ACK rate ratio	0.191600931560364
Idle connection rate ratio	0
Half-open connection rate ratio	0.927765190097077
Packet Service discrete	0.0200121554212495
Bytes Service discrete	0.00925925486408152
Connection Service discrete	0.0028020045725733
Packet Source port discrete	0.0225606502702755
Bytes Source port discrete	0.0146321171288695
Connection Source port discrete	0.0080366050743384
Connection Source port orderedContinuous	0.0155222027424716
Packet TTL discrete	0.114723474946822
Packet TTL orderedContinuous	0.198911719466188
InterPacket delta sortedContinuous	0.5785736634099731
Packet sec orderedContinuous	0.860808954143827
Packet min orderedContinuous	0.693363677630167
Packet GmHour orderedContinuous	0.0344136370430044
Packet LocHour orderedContinuous	0.0955316620021783
Packet weekday orderedContinuous	0.4166666666666667
Bytes sec orderedContinuous	0.867352816369738
Bytes min orderedContinuous	0.563430626340472
Bytes GmHour orderedContinuous	0.0252147247360146
Bytes LocHour orderedContinuous	0.0814568346671174
Continued on next page	

Table L.64 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	0.416666666666667
Packet size orderedContinuous	0.463956877861214
Packets in last w secs orderedContinuous	0.185137872544833
Priv packets time rate sortedContinuous	0.6170375857586708
Unpriv packets time rate sortedContinuous	0.8633236113468209
Connections time rate sortedContinuous	0.3620179199314198
Priv connections connection time rate sortedContinuous	0.7469824823534925
Unpriv connections connection time rate sortedContinuous	0.1821142672219647
Priv packets priv connection time rate sortedContinuous	0.5245101659546307
Unpriv packets unpriv connection time rate sortedContinuous	0.1848530789463916
SYNs connection time rate sortedContinuous	0.4971844240674495
RSTs connection time rate sortedContinuous	0.2403118069811121
FINs connection time rate sortedContinuous	0.2083525234554529
PSH connection time rate sortedContinuous	0.3630456504019996
Establishment errors connection time rate sortedContinuous	0.4534622527330158
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9843275723223688
Ave duration over last w secs sortedContinuous	0.06183714659831176
Priv packets packet rate sortedContinuous	0.6461700895970353
Unpriv packets packet rate sortedContinuous	0.8857461123760438
InterConnection delta sortedContinuous	0.4046190101237623
Connection sec orderedContinuous	0.840866124211211
Connection min orderedContinuous	0.76565406602015
Connection GmHour orderedContinuous	0.0821883136292538
Connection LocHour orderedContinuous	0.0996130332049111
Connection weekday orderedContinuous	0.416666666666667
Connection packet rate sortedContinuous	0.7015298145571451
Connection Priv connections rate sortedContinuous	0.08169069607865958
Connection Unpriv connections rate sortedContinuous	0.1863957850833394
Connection Priv packet rate sortedContinuous	0.1802580342180608
Connection Unpriv packet rate sortedContinuous	0.8308369616975542
Continued on next page	

Table L.64 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.2255712973660916
Connection RSTs rate sortedContinuous	0.2068422085035206
Connection FINs rate sortedContinuous	0.08876316055548048
Connection PSH rate sortedContinuous	0.2909533476468729
Connection Establishment errors rate sortedContinuous	0.7633275099261555
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9995452478399273
Ave duration over last m connections sortedContinuous	0.2527424807566248
Number of packets orderedContinuous	0.0359100872511184
Number of packets in orderedContinuous	0.0417354183257383
Number of packets out orderedContinuous	0.030344379840105
Duration sortedContinuous	0.8233792150948445
Number control packets rate sortedContinuous	0.8410914627647703
Number data packets rate sortedContinuous	0.9055010834200421
Number bytes transferred orderedContinuous	0.0153824843535703
Number bytes transferred in orderedContinuous	0.014079122683534
Number bytes transferred out orderedContinuous	0.0461947835708989
Number data bytes transferred orderedContinuous	0.00995111116029177
Number data bytes transferred in orderedContinuous	0.0101803077658814
Number data bytes transferred out orderedContinuous	0.035659130790205
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.0144021630686445
Max Dst Window orderedContinuous	0.033832946596285
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9570085367406551
Wrong resend rate sortedContinuous	0.9717926014300423
Duplicate ACK rate sortedContinuous	0.9993726474278545
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	0.9993726474278545
Window exceeded rate sortedContinuous	0.9952902280457837
Continued on next page	

Table L.64 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9995164157256378
Number connection errors orderedContinuous	0.164920048302527
Number reset connection orderedContinuous	0.23086407243402
Number other errors orderedContinuous	0.252667251901621
Number disconnection errors orderedContinuous	0.205310896531067
Packet Destination IP nonKeyedSortedContinuous	0.856818628806007
Bytes Destination IP nonKeyedSortedContinuous	0.806135527710291
Connection Destination IP nonKeyedSortedContinuous	0.92493167351588
Packet Source IP nonKeyedSortedContinuous	0.203791948166768
Bytes Source IP nonKeyedSortedContinuous	0.219179416872365
Connection Source IP nonKeyedSortedContinuous	0.615137432135726

Table L.65: Similarity values for individual metrics of basecase 12,
test pair 2

Metric	Similarity
Packets in count	0.000293532201399826
Packets out count	0.000760755176305028
Connections in count	0.00013860101178742
Connections out count	0.000224604566383624
Bytes in count	0.000102197215358646
Bytes out count	0.00060414458178204
SYN-ONLY rate ratio	0
SYN-ACK rate ratio	0
Idle connection rate ratio	1
Half-open connection rate ratio	0
Packet Service discrete	0.000646973155867458
Bytes Service discrete	0.000258579144664012
Connection Service discrete	0.000928973310178138
Continued on next page	

Table L.65 – continued from previous page

Metric	Similarity
Packet Source port discrete	0.000206716547815282
Bytes Source port discrete	0.000159526654496629
Connection Source port discrete	6.80593718518082e-07
Connection Source port orderedContinuous	2.0489958735583e-06
Packet TTL discrete	0.00215201538216709
Packet TTL orderedContinuous	0.0401252459368596
InterPacket delta sortedContinuous	0.2841195400759478
Packet sec orderedContinuous	0.665261314882984
Packet min orderedContinuous	0.54285876731777
Packet GmHour orderedContinuous	0.244584710297593
Packet LocHour orderedContinuous	0.211090607084472
Packet weekday orderedContinuous	0
Bytes sec orderedContinuous	0.667333325354531
Bytes min orderedContinuous	0.513859628713504
Bytes GmHour orderedContinuous	0.194797547601259
Bytes LocHour orderedContinuous	0.174686630873565
Bytes weekday orderedContinuous	0
Packet size orderedContinuous	0.00468627645498649
Packets in last w secs orderedContinuous	2.45326613037173e-05
Priv packets time rate sortedContinuous	0.1647720428261048
Unpriv packets time rate sortedContinuous	0.5681540993214466
Connections time rate sortedContinuous	0.02419549990016723
Priv connections connection time rate sortedContinuous	0.3584307580393309
Unpriv connections connection time rate sortedContinuous	0.1357142857142857
Priv packets priv connection time rate sortedContinuous	0.2026348714795758
Unpriv packets unpriv connection time rate sortedContinuous	0.08410982277740053
SYNs connection time rate sortedContinuous	0.5301235595736393
RSTs connection time rate sortedContinuous	0.2274986972458165
FINs connection time rate sortedContinuous	0.08120071105365221
PSH connection time rate sortedContinuous	0.3312506926942006
Establishment errors connection time rate sortedContinuous	0.9276753702460659
Continued on next page	

Table L.65 – continued from previous page

Metric	Similarity
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	0.9928571428571429
Ave duration over last w secs sortedContinuous	0.05487287052355035
Priv packets packet rate sortedContinuous	0.1308557554124984
Unpriv packets packet rate sortedContinuous	0.683720243167388
InterConnection delta sortedContinuous	0.1091995043833146
Connection sec orderedContinuous	0.633037248901909
Connection min orderedContinuous	0.595785615232199
Connection GmHour orderedContinuous	0.162649223918282
Connection LocHour orderedContinuous	0.131431258108728
Connection weekday orderedContinuous	0
Connection packet rate sortedContinuous	0.5713316016445684
Connection Priv connections rate sortedContinuous	0.02581225727005778
Connection Unpriv connections rate sortedContinuous	0.2123588221980997
Connection Priv packet rate sortedContinuous	0.03194993412384717
Connection Unpriv packet rate sortedContinuous	0.7122935035978514
Connection SYNs rate sortedContinuous	0.06834697874460222
Connection RSTs rate sortedContinuous	0.08701019555778202
Connection FINs rate sortedContinuous	0.9570888222251206
Connection PSH rate sortedContinuous	0.08553609775047868
Connection Establishment errors rate sortedContinuous	0.1445511010728402
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	0.9782608695652174
Ave duration over last m connections sortedContinuous	0.08359324356473803
Number of packets orderedContinuous	0.0106844506140946
Number of packets in orderedContinuous	0.0193245088481609
Number of packets out orderedContinuous	0.0204702386778928
Duration sortedContinuous	0.6958059151796679
Number control packets rate sortedContinuous	0.7173913043478261
Number data packets rate sortedContinuous	0.8823037576234508
Number bytes transferred orderedContinuous	9.0678655367181e-05
Continued on next page	

Table L.65 – continued from previous page

Metric	Similarity
Number bytes transferred in orderedContinuous	0.00158007590629775
Number bytes transferred out orderedContinuous	0.00197850277002975
Number data bytes transferred orderedContinuous	0.000628468567208656
Number data bytes transferred in orderedContinuous	0.000759068195654482
Number data bytes transferred out orderedContinuous	0.0010964226185043
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	0.000117453605825694
Max Dst Window orderedContinuous	4.77478911348082e-05
Urgent rate sortedContinuous	1
Resend rate sortedContinuous	0.9426114426114427
Wrong resend rate sortedContinuous	0.9642857142857143
Duplicate ACK rate sortedContinuous	1
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9285714285714286
Hole rate sortedContinuous	1
Number connection errors orderedContinuous	0.00481828496764844
Number reset connection orderedContinuous	0.0175309147006385
Number other errors orderedContinuous	0.0178471733523801
Number disconnection errors orderedContinuous	0.0141921431282637
Packet Destination IP nonKeyedSortedContinuous	0.269901579672461
Bytes Destination IP nonKeyedSortedContinuous	0.241687148849291
Connection Destination IP nonKeyedSortedContinuous	0.540122898813275
Packet Source IP nonKeyedSortedContinuous	0.643400901595964
Bytes Source IP nonKeyedSortedContinuous	0.450937159928939
Connection Source IP nonKeyedSortedContinuous	0.799975276993596

Table L.66: Similarity values for individual metrics of basecase 12,
test pair 3

Metric	Similarity
Packets in count	0.671737798652236
Packets out count	0.494613167812015
Connections in count	0.176966292134832
Connections out count	0.276700699782035
Bytes in count	0.993464639754901
Bytes out count	0.466246461507291
SYN-ONLY rate ratio	0.705383397132969
SYN-ACK rate ratio	0
Idle connection rate ratio	0
Half-open connection rate ratio	0.702397217775174
Packet Service discrete	0.000936052281838318
Bytes Service discrete	0.000920945902047268
Connection Service discrete	0.00237794534151902
Packet Source port discrete	0.0015126945963254
Bytes Source port discrete	0.00161739089998054
Connection Source port discrete	0.00220475840145154
Connection Source port orderedContinuous	0.00513291557422634
Packet TTL discrete	0.0507784170952856
Packet TTL orderedContinuous	0.0648252822134569
InterPacket delta sortedContinuous	0.2004674678029581
Packet sec orderedContinuous	0.479021588107312
Packet min orderedContinuous	0.685651410681885
Packet GmHour orderedContinuous	0.332558059538861
Packet LocHour orderedContinuous	0.301280128450694
Packet weekday orderedContinuous	0.869073827605485
Bytes sec orderedContinuous	0.592685308135344
Bytes min orderedContinuous	0.692969844270394
Bytes GmHour orderedContinuous	0.341062619297202
Bytes LocHour orderedContinuous	0.313338401044036
Continued on next page	

Table L.66 – continued from previous page

Metric	Similarity
Bytes weekday orderedContinuous	0.897687131023122
Packet size orderedContinuous	0.190940148792667
Packets in last w secs orderedContinuous	0.00894076931611522
Priv packets time rate sortedContinuous	0.871371941141036
Unpriv packets time rate sortedContinuous	0.9329330530423808
Connections time rate sortedContinuous	0.7378806734220848
Priv connections connection time rate sortedContinuous	0.8075618362538007
Unpriv connections connection time rate sortedContinuous	0.7531069835724781
Priv packets priv connection time rate sortedContinuous	0.7627908459155815
Unpriv packets unpriv connection time rate sortedContinuous	0.7149510380970421
SYNs connection time rate sortedContinuous	0.6619116685550387
RSTs connection time rate sortedContinuous	0.4803838205517023
FINs connection time rate sortedContinuous	0.3852091790523799
PSH connection time rate sortedContinuous	0.4575860261354921
Establishment errors connection time rate sortedContinuous	0.6146286666704642
Other errors connection time rate sortedContinuous	1
Disconnection errors connection time rate sortedContinuous	1
Ave duration over last w secs sortedContinuous	0.155313802512919
Priv packets packet rate sortedContinuous	0.7797516336811986
Unpriv packets packet rate sortedContinuous	0.8550386005996104
InterConnection delta sortedContinuous	0.05327653097996169
Connection sec orderedContinuous	0.658023167026536
Connection min orderedContinuous	0.564011549818585
Connection GmHour orderedContinuous	0.331099303053035
Connection LocHour orderedContinuous	0.29715438511841
Connection weekday orderedContinuous	0.940477043382666
Connection packet rate sortedContinuous	0.5970218442384576
Connection Priv connections rate sortedContinuous	0.6778469543380101
Connection Unpriv connections rate sortedContinuous	0.3436905800593403
Connection Priv packet rate sortedContinuous	0.4578413955499301
Connection Unpriv packet rate sortedContinuous	0.4825582178534763

Continued on next page

Table L.66 – continued from previous page

Metric	Similarity
Connection SYNs rate sortedContinuous	0.6233294780289481
Connection RSTs rate sortedContinuous	0.2402742263914115
Connection FINs rate sortedContinuous	0.1612138069390304
Connection PSH rate sortedContinuous	0.1494289269241662
Connection Establishment errors rate sortedContinuous	0.6491969752712402
Connection Other errors rate sortedContinuous	1
Connection Disconnection errors rate sortedContinuous	1
Ave duration over last m connections sortedContinuous	0.001940118758771735
Number of packets orderedContinuous	0.228397266829738
Number of packets in orderedContinuous	0.256800790692446
Number of packets out orderedContinuous	0.192854242036559
Duration sortedContinuous	0.3891242242828489
Number control packets rate sortedContinuous	0.3721832721458359
Number data packets rate sortedContinuous	0.3415455932361868
Number bytes transferred orderedContinuous	0.0348437921798032
Number bytes transferred in orderedContinuous	0.0254875322236684
Number bytes transferred out orderedContinuous	0.0352367271711361
Number data bytes transferred orderedContinuous	0.0130886479189069
Number data bytes transferred in orderedContinuous	0.00912724170074727
Number data bytes transferred out orderedContinuous	0.0321285194376414
Fragmented packets rate sortedContinuous	1
Bad fragment rate sortedContinuous	1
Max Src Window orderedContinuous	7.89785187129295e-05
Max Dst Window orderedContinuous	0.022210670375269
Urgent rate sortedContinuous	0.9989539748953975
Resend rate sortedContinuous	0.9888143552767484
Wrong resend rate sortedContinuous	0.997007320088493
Duplicate ACK rate sortedContinuous	0.9989539748953975
Wrong ACK sortedContinuous	1
Wrong data packet size rate sortedContinuous	1
Window exceeded rate sortedContinuous	0.9781918015275555
Continued on next page	

Table L.66 – continued from previous page

Metric	Similarity
Hole rate sortedContinuous	0.9967747559274756
Number connection errors orderedContinuous	0.0870369412991999
Number reset connection orderedContinuous	0.246491291265687
Number other errors orderedContinuous	0.250976338255498
Number disconnection errors orderedContinuous	0.250976338255498
Packet Destination IP nonKeyedSortedContinuous	0.221518126552476
Bytes Destination IP nonKeyedSortedContinuous	0.0994290713515927
Connection Destination IP nonKeyedSortedContinuous	0.477910217075628
Packet Source IP nonKeyedSortedContinuous	0.484702281948999
Bytes Source IP nonKeyedSortedContinuous	0.214207966360144
Connection Source IP nonKeyedSortedContinuous	0.368561073131439

Appendix M

Analysis Details

This appendix contains additional tables and figures for chapter 9 which were not directly necessary for the accompanying discussion in that chapter.

Table M.1: Weights for the real data basecases sorted from highest to lowest by absolute value.

Metric	Weight
FINs connection time rate sortedContinuous	-0.5003
Unpriv connections connection time rate sortedContinuous	0.3721
Packet Destination IP nonKeyedSortedContinuous	0.2808
InterPacket delta sortedContinuous	-0.2662
InterConnection delta sortedContinuous	0.2474
Packet TTL discrete	0.231
Priv packets time rate sortedContinuous	-0.2287
Priv connections connection time rate sortedContinuous	-0.2282
Unpriv packets time rate sortedContinuous	0.2238
Connection PSH rate sortedContinuous	0.2164
Connection Priv packet rate sortedContinuous	0.2151
Connection FINs rate sortedContinuous	0.2143
Resend rate sortedContinuous	-0.2141
Connections out count	-0.2127
Ave duration over last w secs sortedContinuous	0.2015
Continued on next page	

Table M.1 – continued from previous page

Metric	Weight
Ave duration over last m connections sortedContinuous	-0.1978
RSTs connection time rate sortedContinuous	0.1971
Connections time rate sortedContinuous	-0.1963
Connection Destination IP nonKeyedSortedContinuous	-0.1925
Number other errors orderedContinuous	0.1915
Number bytes transferred in orderedContinuous	0.1839
Connection Establishment errors rate sortedContinuous	0.1781
Half-open connection rate ratio	-0.1756
Priv packets packet rate sortedContinuous	-0.1692
Packets in count	0.1625
Connection RSTs rate sortedContinuous	0.1534
Connection weekday orderedContinuous	-0.153
Connection Unpriv packet rate sortedContinuous	-0.1526
Number of packets in orderedContinuous	-0.1496
Connection Unpriv connections rate sortedContinuous	0.1462
Unpriv packets unpriv connection time rate sortedContinuous	0.1393
SYN-ACK rate ratio	-0.1364
Connection min orderedContinuous	-0.134
Connection packet rate sortedContinuous	-0.129
Number data bytes transferred in orderedContinuous	-0.1259
Bytes in count	0.1185
Duration sortedContinuous	0.1128
Connection sec orderedContinuous	-0.1092
Establishment errors connection time rate sortedContinuous	-0.1039
Packet weekday orderedContinuous	-0.1025
Number connection errors orderedContinuous	0.1001
Connection Source IP nonKeyedSortedContinuous	-0.0958
Number data packets rate sortedContinuous	0.094
Unpriv packets packet rate sortedContinuous	0.092
Bytes weekday orderedContinuous	-0.0748
Packet Service discrete	-0.0716
Continued on next page	

Table M.1 – continued from previous page

Metric	Weight
Connection Service discrete	-0.0383
Packets in last w secs orderedContinuous	-0.0309
Packet Source port discrete	-0.0273
Connection Source port discrete	-0.0199
Connection Source port orderedContinuous	0.0113

M.1 Reports

M.1.1 Report for basecase 1, testing pair 2

Similarity report between sigcomm2004-98.93.251-20040831-1900
and sigcomm2004-98.93.251-20040831-2000

Overall scaled similarity: 0.81401262523546

Goal was 0.825

This value is based on the following characteristics, ordered from highest to lowest impact. The number following the characteristic is the actual impact that particular characteristic contributed. Negative numbers mean that characteristic had a negative impact; that is, the characteristic indicates that the networks are less similar.

```

FINs connection time rate sortedContinuous: -0.352698838509653
Unpriv connections connection time rate sortedContinuous: 0.260283999389003
InterPacket delta sortedContinuous: -0.247358553494784
Packet Destination IP nonKeyedSortedContinuous: 0.234140865906291
InterConnection delta sortedContinuous: 0.212635488047483
Resend rate sortedContinuous: -0.186984055702948
Priv connections connection time rate sortedContinuous: -0.184902310701664
Number other errors orderedContinuous: 0.172416824545301
Half-open connection rate ratio: -0.169954853106609
Connection FINs rate sortedContinuous: 0.16923984079938

```

Connection Destination IP nonKeyedSortedContinuous: -0.168871285020834
Unpriv packets time rate sortedContinuous: 0.168687455102995
Connection weekday orderedContinuous: -0.153
Connection RSTs rate sortedContinuous: 0.149670872791799
Packets in count: 0.14441675321225
Connections out count: -0.134540580716866
Connection Establishment errors rate sortedContinuous: 0.134209298259333
RSTs connection time rate sortedContinuous: 0.123202815338198
Priv packets time rate sortedContinuous: -0.117874369244019
Connections time rate sortedContinuous: -0.117330102891174
Bytes in count: 0.113834941148787
Duration sortedContinuous: 0.11069179789968
Connection sec orderedContinuous: -0.102872318125177
Packet weekday orderedContinuous: -0.1025
Ave duration over last m connections sortedContinuous: -0.0984774040342231
Connection Unpriv packet rate sortedContinuous: -0.0966323593963233
Number data packets rate sortedContinuous: 0.0930083660767013
Connection PSH rate sortedContinuous: 0.0915595050310237
Unpriv packets unpriv connection time rate sortedContinuous: 0.0852085473696129
Priv packets packet rate sortedContinuous: -0.0768826695169646
Bytes weekday orderedContinuous: -0.0748
Connection Unpriv connections rate sortedContinuous: 0.0728944084016019
Connection Source IP nonKeyedSortedContinuous: -0.070538419356642
Unpriv packets packet rate sortedContinuous: 0.0672683687302999
Number connection errors orderedContinuous: 0.0663570500056252
Connection min orderedContinuous: -0.0604526257789789
Packet TTL discrete: 0.0601717875861952
Connection packet rate sortedContinuous: -0.0498103491693571
Number of packets in orderedContinuous: -0.0462018472980632
Connection Priv packet rate sortedContinuous: 0.0446598797810585
Ave duration over last w secs sortedContinuous: 0.0429599034959817
SYN-ACK rate ratio: -0.0318332760671336
Establishment errors connection time rate sortedContinuous: -0.027963720614672
Number bytes transferred in orderedContinuous: 0.0247921282802081
Packets in last w secs orderedContinuous: -0.0162276422435271

```
Number data bytes transferred in orderedContinuous: -0.0116755839791505
Packet Service discrete: -0.00150487207470761
Connection Service discrete: -0.000822336707763506
Packet Source port discrete: -0.000606801836614515
Connection Source port orderedContinuous: 0.000378590899950151
Connection Source port discrete: -0.000359687275447825
```

The following characteristics did not affect the outcome of the scaled similarity calculation:

```
Number data bytes transferred orderedContinuous
Connection Priv connections rate sortedContinuous
Bytes LocHour orderedContinuous
Number disconnection errors orderedContinuous
Number of packets out orderedContinuous
Packet Source IP nonKeyedSortedContinuous
Urgent rate sortedContinuous
Wrong ACK sortedContinuous
Packet size orderedContinuous
Bytes GmHour orderedContinuous
Connections in count
Bytes min orderedContinuous
Duplicate ACK rate sortedContinuous
Number data bytes transferred out orderedContinuous
Connection Other errors rate sortedContinuous
Idle connection rate ratio
Connection Disconnection errors rate sortedContinuous
Packet TTL orderedContinuous
Bytes sec orderedContinuous
Bytes Service discrete
Packet GmHour orderedContinuous
Bytes Destination IP nonKeyedSortedContinuous
Packet sec orderedContinuous
Number control packets rate sortedContinuous
Bad fragment rate sortedContinuous
Connection LocHour orderedContinuous
```

Wrong data packet size rate sortedContinuous
Priv packets priv connection time rate sortedContinuous
Max Dst Window orderedContinuous
Max Src Window orderedContinuous
Number reset connection orderedContinuous
Other errors connection time rate sortedContinuous
Number bytes transferred out orderedContinuous
PSH connection time rate sortedContinuous
Window exceeded rate sortedContinuous
Fragmented packets rate sortedContinuous
Bytes out count
Disconnection errors connection time rate sortedContinuous
Packets out count
SYNs connection time rate sortedContinuous
Packet LocHour orderedContinuous
Wrong resend rate sortedContinuous
Bytes Source port discrete
Connection SYNs rate sortedContinuous
Packet min orderedContinuous
Number of packets orderedContinuous
Bytes Source IP nonKeyedSortedContinuous
SYN-ONLY rate ratio
Connection GmHour orderedContinuous
Hole rate sortedContinuous
Number bytes transferred orderedContinuous

M.1.2 Report for basecase 3, testing pair 2

Simularity report between lbl-128.3.23-20041216-2117
and lbl-128.3.23-20050107-2326

Overall scaled simularity: 0.847397653910108
Goal was 0.825

This value is based on the following characteristics, ordered from highest to lowest impact. The number following the characteristic is the actual impact that particular characteristic contributed. Negative numbers mean that characteristic had a negative impact; that is, the characteristic indicates that the networks are less similar.

```

FINs connection time rate sortedContinuous: -0.451726164542137
Unpriv connections connection time rate sortedContinuous: 0.341201467901922
Packet Destination IP nonKeyedSortedContinuous: 0.246025972751202
Priv connections connection time rate sortedContinuous: -0.2224858250164
Priv packets time rate sortedContinuous: -0.211559141996489
Unpriv packets time rate sortedContinuous: 0.205513826852744
InterPacket delta sortedContinuous: -0.1828635481945
Connection Destination IP nonKeyedSortedContinuous: -0.176351747779325
Connection Priv packet rate sortedContinuous: 0.160548338391932
Connection FINs rate sortedContinuous: 0.155452403822209
Priv packets packet rate sortedContinuous: -0.154893374202076
Connection PSH rate sortedContinuous: 0.152950590729333
Resend rate sortedContinuous: -0.152035115531969
Connections time rate sortedContinuous: -0.150857324736036
RSTs connection time rate sortedContinuous: 0.150613030031108
Ave duration over last w secs sortedContinuous: 0.132150539087576
Connection Unpriv connections rate sortedContinuous: 0.13162212809733
Connection Establishment errors rate sortedContinuous: 0.12931521436769
Connection RSTs rate sortedContinuous: 0.12652522933718
InterConnection delta sortedContinuous: 0.125090561824583
Half-open connection rate ratio: -0.116928992837367
Unpriv packets unpriv connection time rate sortedContinuous: 0.114138258786502

```

Connection packet rate sortedContinuous: -0.108337117214169
 Duration sortedContinuous: 0.106906035963473
 Ave duration over last m connections sortedContinuous: -0.102932105099198
 Connection Unpriv packet rate sortedContinuous: -0.0947540711197375
 Number data packets rate sortedContinuous: 0.0910998722554127
 Connections out count: -0.0879610669693531
 Packet TTL discrete: 0.0877127147913761
 Unpriv packets packet rate sortedContinuous: 0.0837124424418048
 SYN-ACK rate ratio: -0.0831648917495613
 Connection sec orderedContinuous: -0.081462197996426
 Packets in count: 0.0794924328398291
 Connection min orderedContinuous: -0.0790996170211541
 Connection weekday orderedContinuous: -0.0593841510996489
 Establishment errors connection time rate sortedContinuous: -0.0536957502436637
 Connection Source IP nonKeyedSortedContinuous: -0.0451769241221859
 Number other errors orderedContinuous: 0.040863655279126
 Packet weekday orderedContinuous: -0.036202170755618
 Number of packets in orderedContinuous: -0.029059258298152
 Bytes weekday orderedContinuous: -0.0259649205722373
 Number bytes transferred in orderedContinuous: 0.0236497639782487
 Packets in last w secs orderedContinuous: -0.0188243604620002
 Number connection errors orderedContinuous: 0.0166727040348088
 Bytes in count: 0.012622043577187
 Number data bytes transferred in orderedContinuous: -0.00925225553054042
 Packet Service discrete: -0.00290285107364261
 Connection Service discrete: -0.00241227702675782
 Packet Source port discrete: -0.00105040766148412
 Connection Source port discrete: -0.000634324051229365
 Connection Source port orderedContinuous: 0.000490379670588931

The following characteristics did not affect the outcome of the scaled similarity calculation:

Number data bytes transferred orderedContinuous
 Connection Priv connections rate sortedContinuous
 Bytes LocHour orderedContinuous

Number disconnection errors orderedContinuous
Number of packets out orderedContinuous
Packet Source IP nonKeyedSortedContinuous
Urgent rate sortedContinuous
Wrong ACK sortedContinuous
Packet size orderedContinuous
Bytes GmHour orderedContinuous
Connections in count
Bytes min orderedContinuous
Duplicate ACK rate sortedContinuous
Number data bytes transferred out orderedContinuous
Connection Other errors rate sortedContinuous
Idle connection rate ratio
Connection Disconnection errors rate sortedContinuous
Packet TTL orderedContinuous
Bytes sec orderedContinuous
Bytes Service discrete
Packet GmHour orderedContinuous
Bytes Destination IP nonKeyedSortedContinuous
Packet sec orderedContinuous
Number control packets rate sortedContinuous
Bad fragment rate sortedContinuous
Connection LocHour orderedContinuous
Wrong data packet size rate sortedContinuous
Priv packets priv connection time rate sortedContinuous
Max Dst Window orderedContinuous
Max Src Window orderedContinuous
Number reset connection orderedContinuous
Other errors connection time rate sortedContinuous
Number bytes transferred out orderedContinuous
PSH connection time rate sortedContinuous
Window exceeded rate sortedContinuous
Fragmented packets rate sortedContinuous
Bytes out count
Disconnection errors connection time rate sortedContinuous

Packets out count
SYNs connection time rate sortedContinuous
Packet LocHour orderedContinuous
Wrong resend rate sortedContinuous
Bytes Source port discrete
Connection SYNs rate sortedContinuous
Packet min orderedContinuous
Number of packets orderedContinuous
Bytes Source IP nonKeyedSortedContinuous
SYN-ONLY rate ratio
Connection GmHour orderedContinuous
Hole rate sortedContinuous
Number bytes transferred orderedContinuous

M.1.3 Report for basecase 3, testing pair 3

Simularity report between lbl-131.243.155-20041216-2117
and lbl-131.243.155-20050106-2225

Overall scaled simularity: 0.322112149998882

Goal was 0.825

This value is based on the following characteristics, ordered from highest to lowest impact. The number following the characteristic is the actual impact that particular characteristic contributed. Negative numbers mean that characteristic had a negative impact; that is, the characteristic indicates that the networks are less similar.

```

FINs connection time rate sortedContinuous: -0.386832223004541
Packet Destination IP nonKeyedSortedContinuous: 0.250705127316763
Unpriv connections connection time rate sortedContinuous: 0.244843842309552
Priv connections connection time rate sortedContinuous: -0.223531870351758
Resend rate sortedContinuous: -0.211310245046027
Connections out count: -0.207437973174367
Unpriv packets time rate sortedContinuous: 0.200945788657577
Priv packets time rate sortedContinuous: -0.199595498806374
Connection Destination IP nonKeyedSortedContinuous: -0.183953025644273
Connection Establishment errors rate sortedContinuous: 0.171592588247732
Number other errors orderedContinuous: 0.170808949576148
InterConnection delta sortedContinuous: 0.160198333540255
Connection FINs rate sortedContinuous: 0.159484720087944
Connection weekday orderedContinuous: -0.153
Half-open connection rate ratio: -0.143194065342688
RSTs connection time rate sortedContinuous: 0.138207068851989
Priv packets packet rate sortedContinuous: -0.137955567469758
InterPacket delta sortedContinuous: -0.136744062553104
SYN-ACK rate ratio: -0.1364
Ave duration over last w secs sortedContinuous: 0.128158287672153
Connection min orderedContinuous: -0.125531638063745
Ave duration over last m connections sortedContinuous: -0.120234274395009

```

Connection RSTs rate sortedContinuous: 0.120123382841722
 Duration sortedContinuous: 0.107292100322708
 Packet weekday orderedContinuous: -0.1025
 Establishment errors connection time rate sortedContinuous: -0.0927046867451141
 Number data packets rate sortedContinuous: 0.0926870406470185
 Connection Unpriv connections rate sortedContinuous: 0.0916225833431893
 Number connection errors orderedContinuous: 0.0893568052872655
 Connection Unpriv packet rate sortedContinuous: -0.0865395502811551
 Unpriv packets packet rate sortedContinuous: 0.0781857103934612
 Connection Source IP nonKeyedSortedContinuous: -0.0781821078769616
 Bytes weekday orderedContinuous: -0.0748
 Connection sec orderedContinuous: -0.0565269660822016
 Number of packets in orderedContinuous: -0.0519129742681424
 Unpriv packets unpriv connection time rate sortedContinuous: 0.0469151532813261
 Packet TTL discrete: 0.0448489485649137
 Connections time rate sortedContinuous: -0.0359208296916781
 Bytes in count: 0.0335490444593125
 Connection Priv packet rate sortedContinuous: 0.0319916371365682
 Packets in count: 0.0290310849299482
 Number bytes transfered in orderedContinuous: 0.0197709440075895
 Connection PSH rate sortedContinuous: 0.0166956026857
 Connection packet rate sortedContinuous: -0.0160684256709555
 Number data bytes transfered in orderedContinuous: -0.0101813404296085
 Packets in last w secs orderedContinuous: -0.00371929469555681
 Connection Service discrete: -0.00218709973947789
 Packet Service discrete: -0.00196201977593952
 Connection Source port discrete: -0.00182065164760848
 Connection Source port orderedContinuous: 0.00162527816714698
 Packet Source port discrete: -0.000781481573058961

The following characteristics did not affect the outcome of the scaled similarity calculation:

Number data bytes transfered orderedContinuous
 Connection Priv connections rate sortedContinuous
 Bytes LocHour orderedContinuous

Number disconnection errors orderedContinuous
Number of packets out orderedContinuous
Packet Source IP nonKeyedSortedContinuous
Urgent rate sortedContinuous
Wrong ACK sortedContinuous
Packet size orderedContinuous
Bytes GmHour orderedContinuous
Connections in count
Bytes min orderedContinuous
Duplicate ACK rate sortedContinuous
Number data bytes transferred out orderedContinuous
Connection Other errors rate sortedContinuous
Idle connection rate ratio
Connection Disconnection errors rate sortedContinuous
Packet TTL orderedContinuous
Bytes sec orderedContinuous
Bytes Service discrete
Packet GmHour orderedContinuous
Bytes Destination IP nonKeyedSortedContinuous
Packet sec orderedContinuous
Number control packets rate sortedContinuous
Bad fragment rate sortedContinuous
Connection LocHour orderedContinuous
Wrong data packet size rate sortedContinuous
Priv packets priv connection time rate sortedContinuous
Max Dst Window orderedContinuous
Max Src Window orderedContinuous
Number reset connection orderedContinuous
Other errors connection time rate sortedContinuous
Number bytes transferred out orderedContinuous
PSH connection time rate sortedContinuous
Window exceeded rate sortedContinuous
Fragmented packets rate sortedContinuous
Bytes out count
Disconnection errors connection time rate sortedContinuous

Packets out count
SYNs connection time rate sortedContinuous
Packet LocHour orderedContinuous
Wrong resend rate sortedContinuous
Bytes Source port discrete
Connection SYNs rate sortedContinuous
Packet min orderedContinuous
Number of packets orderedContinuous
Bytes Source IP nonKeyedSortedContinuous
SYN-ONLY rate ratio
Connection GmHour orderedContinuous
Hole rate sortedContinuous
Number bytes transferred orderedContinuous

M.1.4 Report for basecase 8, testing pair 3

Simularity report between dsl1-69.225.88.159-20070302-1800
and sotm27-172.16.134-20030302-1800

Overall scaled simularity: 0.825898599300724

Goal was 0.175

This value is based on the following characteristics, ordered from highest to lowest impact. The number following the characteristic is the actual impact that particular characteristic contributed. Negative numbers mean that characteristic had a negative impact; that is, the characteristic indicates that the networks are less similar.

```

FINs connection time rate sortedContinuous: -0.255255102040816
Unpriv connections connection time rate sortedContinuous: 0.169090340136054
Connection Establishment errors rate sortedContinuous: 0.155774074074074
Packet Destination IP nonKeyedSortedContinuous: 0.150332331564632
RSTs connection time rate sortedContinuous: 0.145813775510204
Resend rate sortedContinuous: -0.142733333333333
SYN-ACK rate ratio: -0.1364
Connection Priv packet rate sortedContinuous: 0.132068518343312
Unpriv packets time rate sortedContinuous: 0.121023831428762
Establishment errors connection time rate sortedContinuous: -0.101143469387755
Connection Destination IP nonKeyedSortedContinuous: -0.0992769655620533
Unpriv packets packet rate sortedContinuous: 0.0820125620251256
Connection packet rate sortedContinuous: -0.0809635206433292
Priv packets packet rate sortedContinuous: -0.0808539023294043
Connection min orderedContinuous: -0.0749121475085345
Connection Source IP nonKeyedSortedContinuous: -0.0744182980810754
Connection sec orderedContinuous: -0.0724193445175386
Priv packets time rate sortedContinuous: -0.0708441644772198
Unpriv packets unpriv connection time rate sortedContinuous: 0.0630561507936508
Bytes in count: 0.0592582906716976
Number data packets rate sortedContinuous: 0.0562690830539144
Priv connections connection time rate sortedContinuous: -0.055967029478458

```

Connection RSTs rate sortedContinuous: 0.0527066666666667
Packets in count: 0.0496115173674588
InterConnection delta sortedContinuous: 0.0454837211498856
Connection Unpriv packet rate sortedContinuous: -0.04545494103749
Connections time rate sortedContinuous: -0.0451117684960728
Number connection errors orderedContinuous: 0.0393607251468808
Half-open connection rate ratio: -0.0390222222222222
Connections out count: -0.0352474285714286
InterPacket delta sortedContinuous: -0.0244725489228299
Connection FINs rate sortedContinuous: 0.0219794871794872
Duration sortedContinuous: 0.0162224288929042
Ave duration over last w secs sortedContinuous: 0.011331260118528
Connection Unpriv connections rate sortedContinuous: 0.00960384613142997
Connection PSH rate sortedContinuous: 0.00887794871794872
Number of packets in orderedContinuous: -0.00543857568509681
Packet TTL discrete: 0.00252538341407881
Ave duration over last m connections sortedContinuous: -0.00156420592589532
Number bytes transferred in orderedContinuous: 0.000735084407058287
Packet Service discrete: -0.000364913703194408
Packets in last w secs orderedContinuous: -0.00023029833582983
Number data bytes transferred in orderedContinuous: -7.81818417556233e-05
Connection Service discrete: -6.56545501149032e-05
Connection Source port orderedContinuous: 1.12705472592468e-05
Connection Source port discrete: -1.12680425937977e-05
Packet Source port discrete: -4.13346247620325e-07
Connection weekday orderedContinuous: 0
Number other errors orderedContinuous: 0
Packet weekday orderedContinuous: 0
Bytes weekday orderedContinuous: 0

The following characteristics did not affect the outcome of the scaled similarity calculation:

Number data bytes transferred orderedContinuous
Connection Priv connections rate sortedContinuous
Bytes LocHour orderedContinuous

Number disconnection errors orderedContinuous
Number of packets out orderedContinuous
Packet Source IP nonKeyedSortedContinuous
Urgent rate sortedContinuous
Wrong ACK sortedContinuous
Packet size orderedContinuous
Bytes GmHour orderedContinuous
Connections in count
Bytes min orderedContinuous
Duplicate ACK rate sortedContinuous
Number data bytes transferred out orderedContinuous
Connection Other errors rate sortedContinuous
Idle connection rate ratio
Connection Disconnection errors rate sortedContinuous
Packet TTL orderedContinuous
Bytes sec orderedContinuous
Bytes Service discrete
Packet GmHour orderedContinuous
Bytes Destination IP nonKeyedSortedContinuous
Packet sec orderedContinuous
Number control packets rate sortedContinuous
Bad fragment rate sortedContinuous
Connection LocHour orderedContinuous
Wrong data packet size rate sortedContinuous
Priv packets priv connection time rate sortedContinuous
Max Dst Window orderedContinuous
Max Src Window orderedContinuous
Number reset connection orderedContinuous
Other errors connection time rate sortedContinuous
Number bytes transferred out orderedContinuous
PSH connection time rate sortedContinuous
Window exceeded rate sortedContinuous
Fragmented packets rate sortedContinuous
Bytes out count
Disconnection errors connection time rate sortedContinuous

Packets out count
SYNs connection time rate sortedContinuous
Packet LocHour orderedContinuous
Wrong resend rate sortedContinuous
Bytes Source port discrete
Connection SYNs rate sortedContinuous
Packet min orderedContinuous
Number of packets orderedContinuous
Bytes Source IP nonKeyedSortedContinuous
SYN-ONLY rate ratio
Connection GmHour orderedContinuous
Hole rate sortedContinuous
Number bytes transferred orderedContinuous

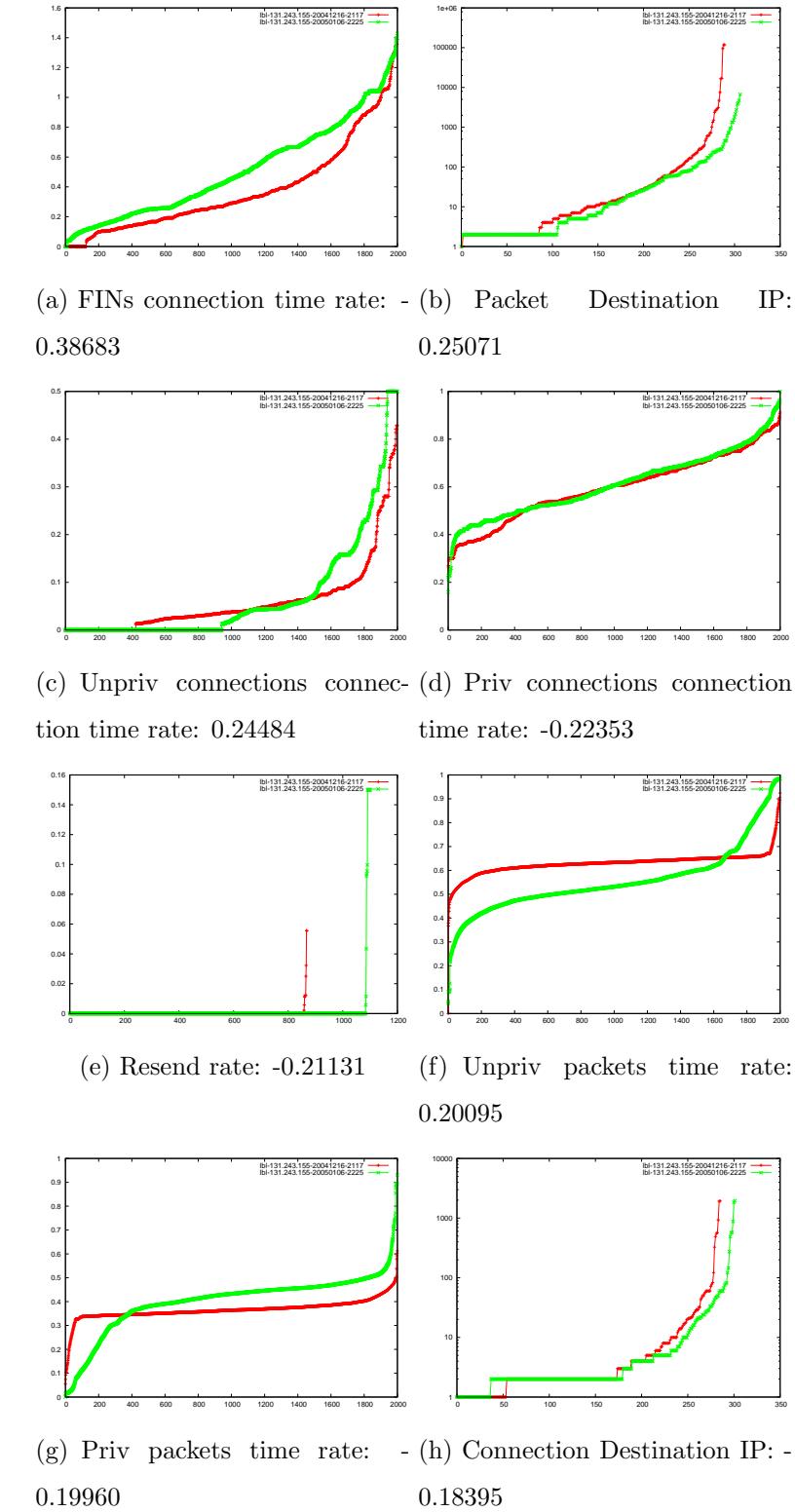
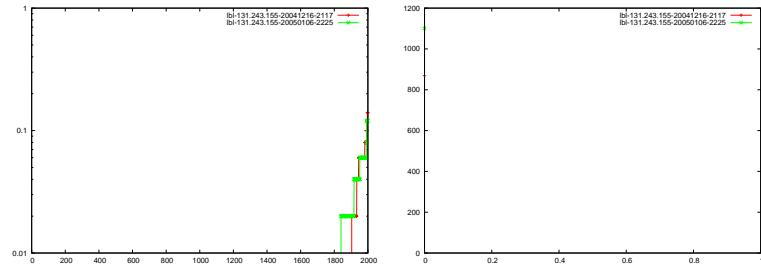
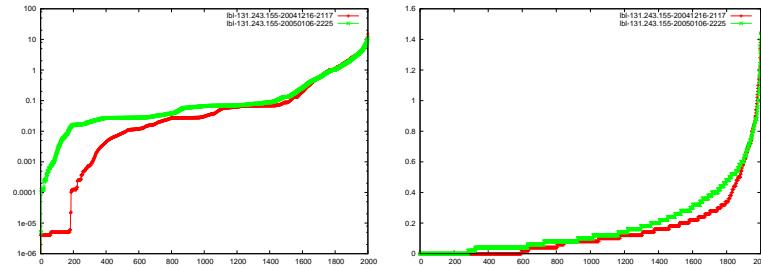


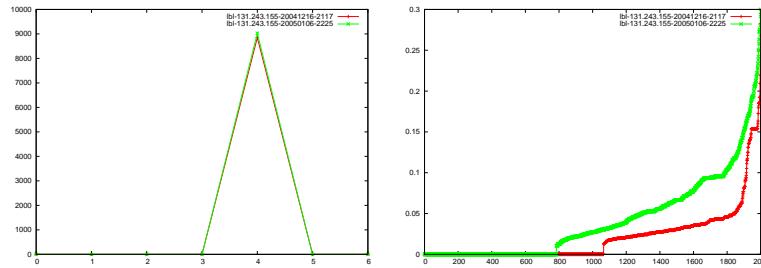
Figure M.1: Plots of characteristics 1 through 8 from basecase 3, pair 3



(a) Connection Establishment errors rate: 0.17159
(b) Number other errors: 0.17081



(c) InterConnection delta: 0.16020
(d) Connection FINs rate: 0.15948



(e) Connection weekday: 0.15300
(f) RSTs connection time rate: 0.13821

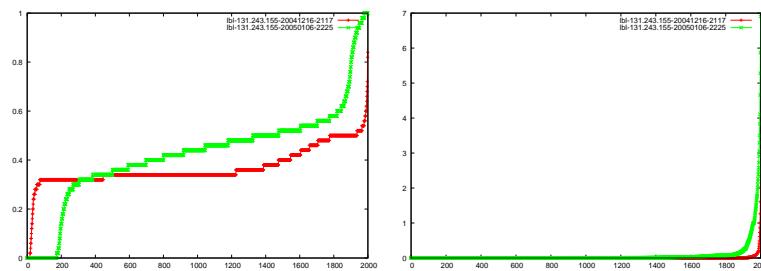


Figure M.2: Plots of characteristics 9 through 16 from basecase 3, pair 3

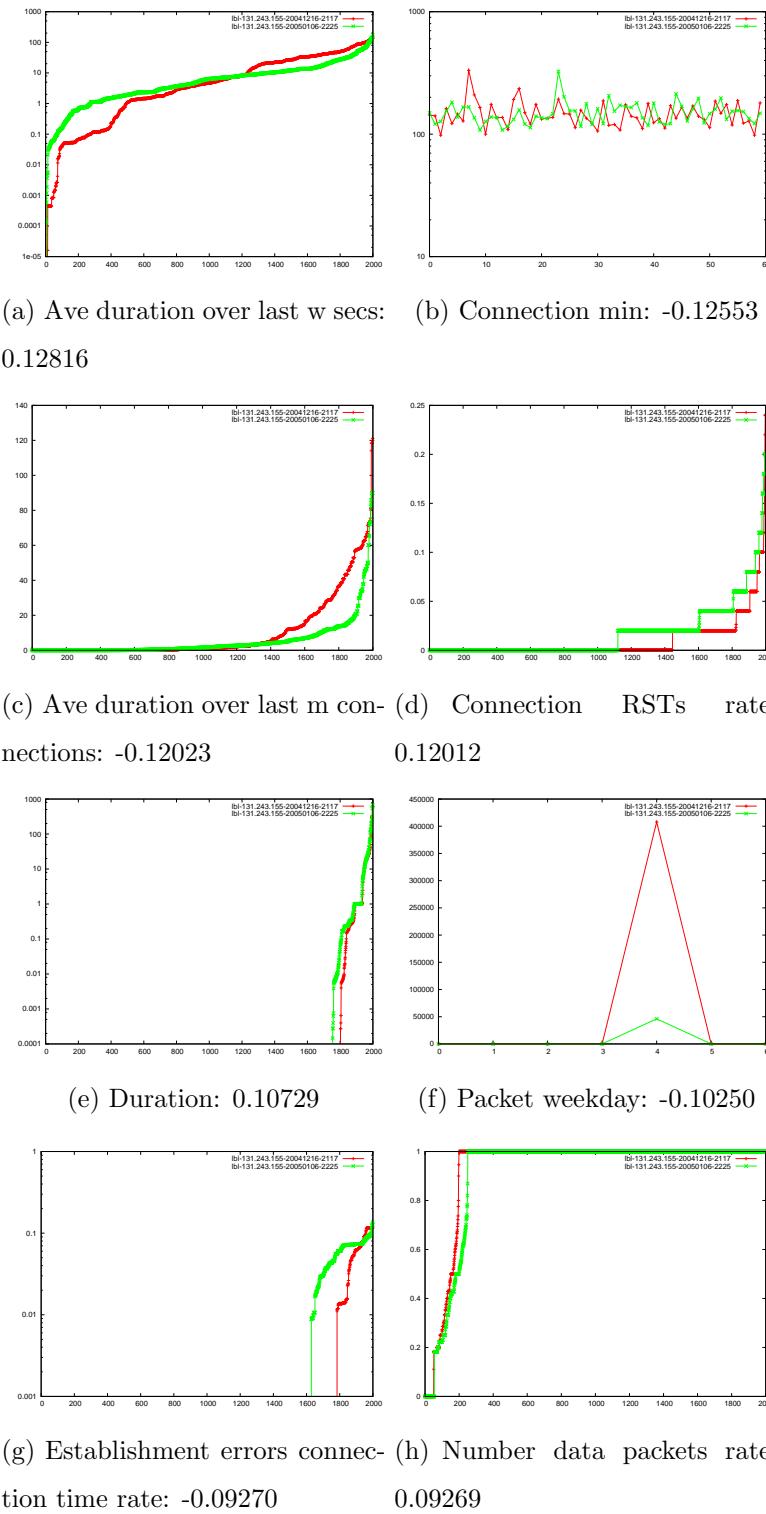


Figure M.3: Plots of characteristics 17 through 24 from basecase 3, pair 3

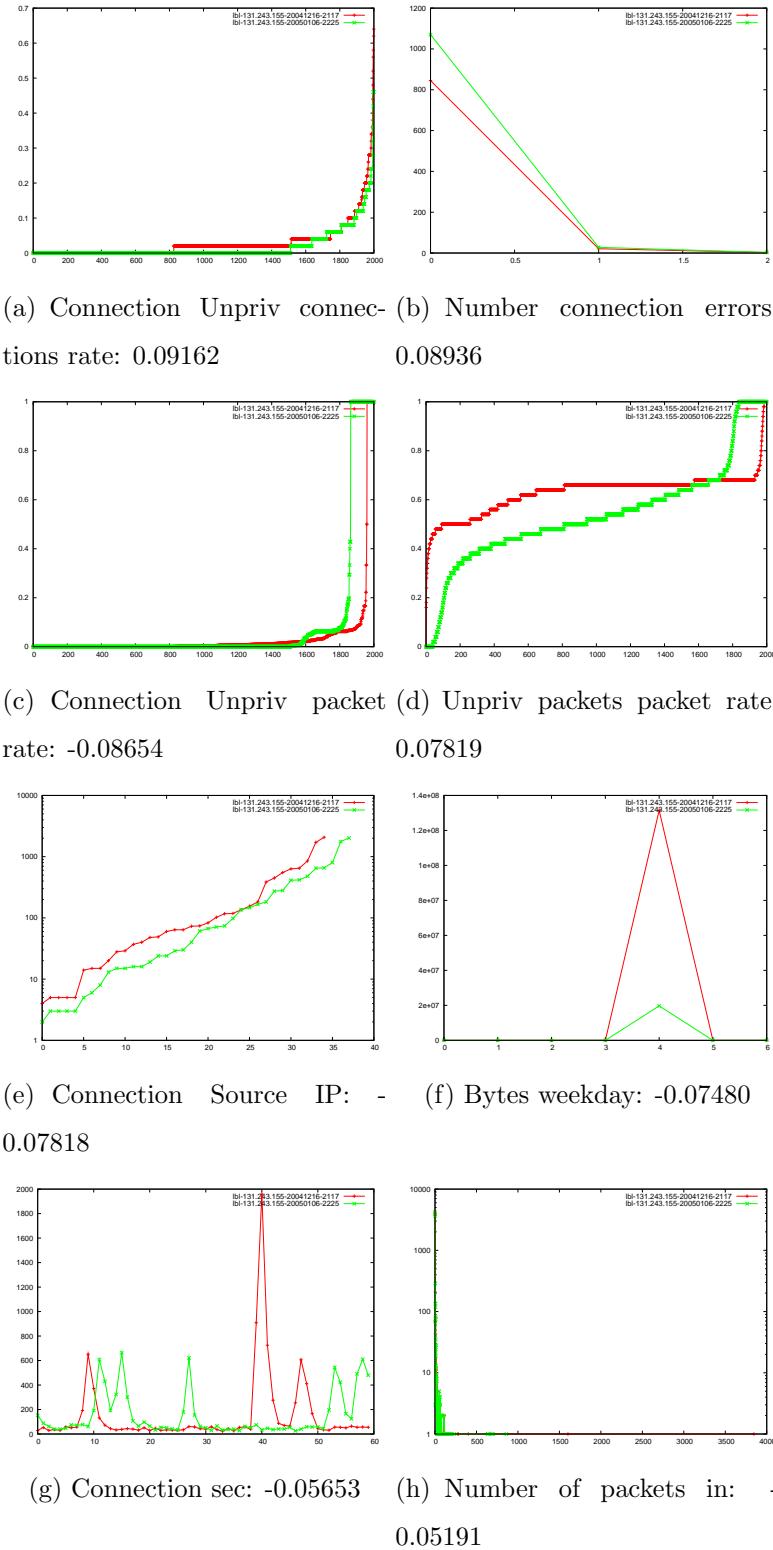


Figure M.4: Plots of characteristics 25 through 32 from basecase 3, pair 3

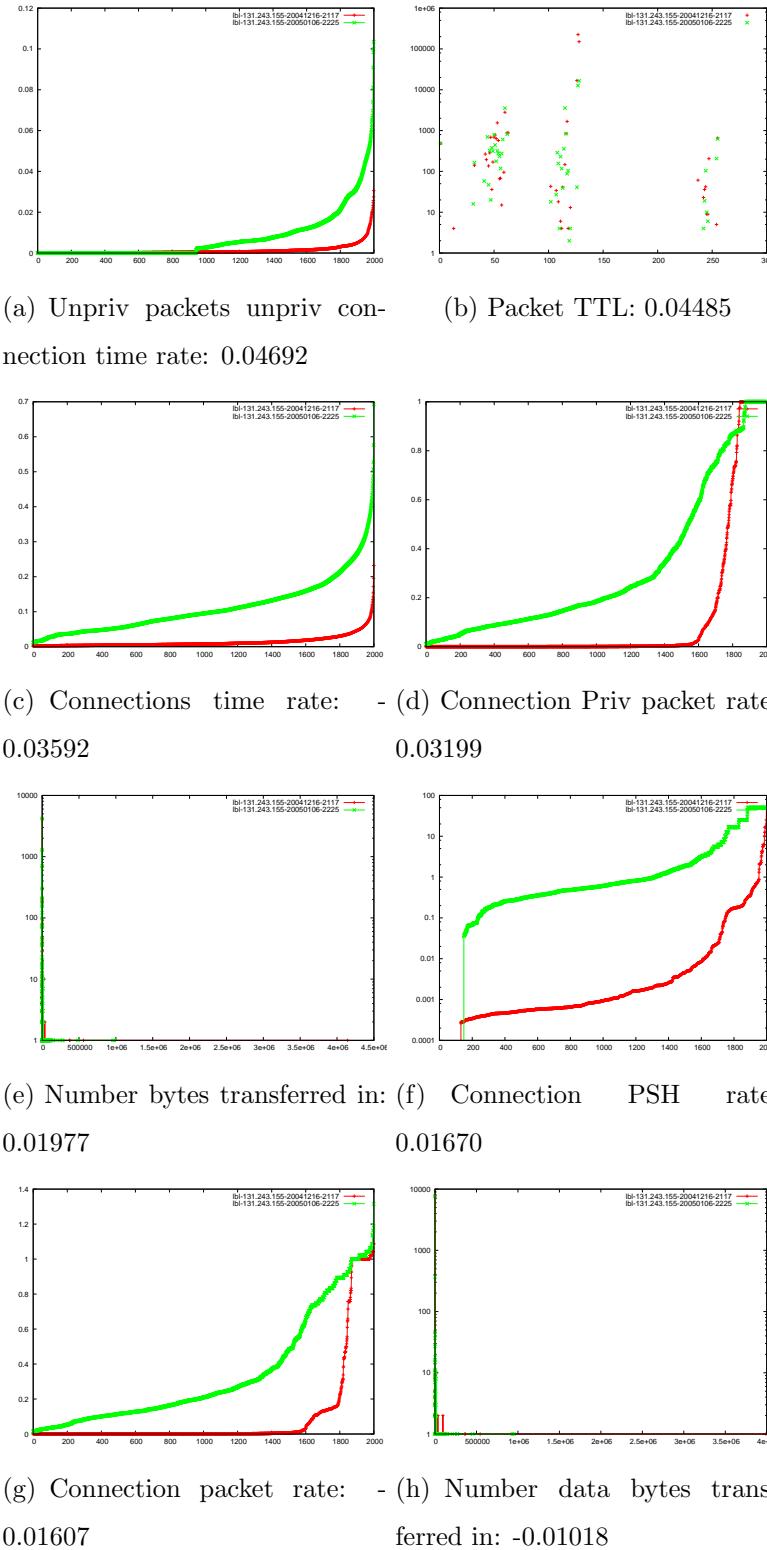


Figure M.5: Plots of characteristics 33 through 40 from basecase 3, pair 3

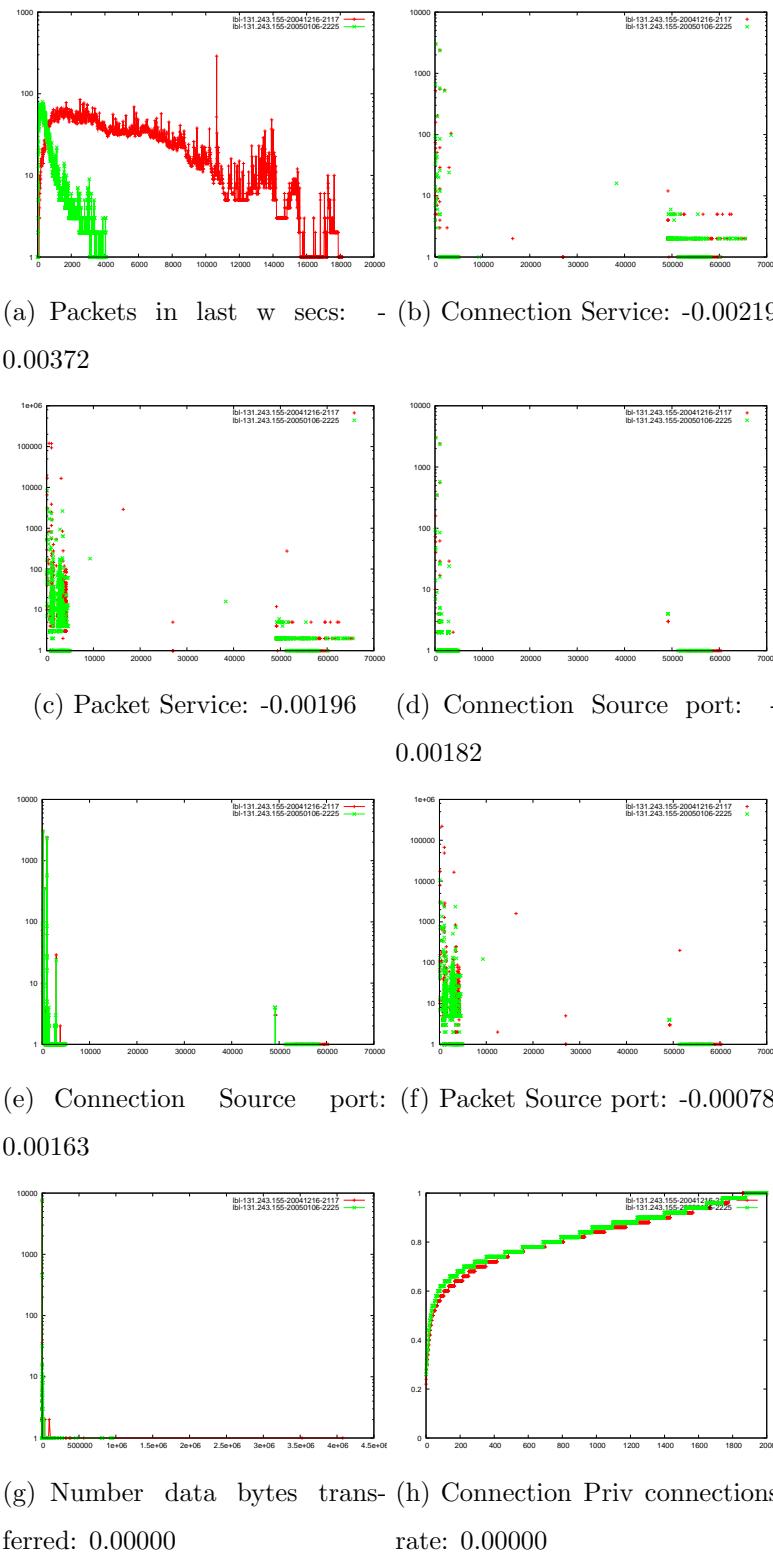


Figure M.6: Plots of characteristics 41 through 48 from basecase 3, pair 3

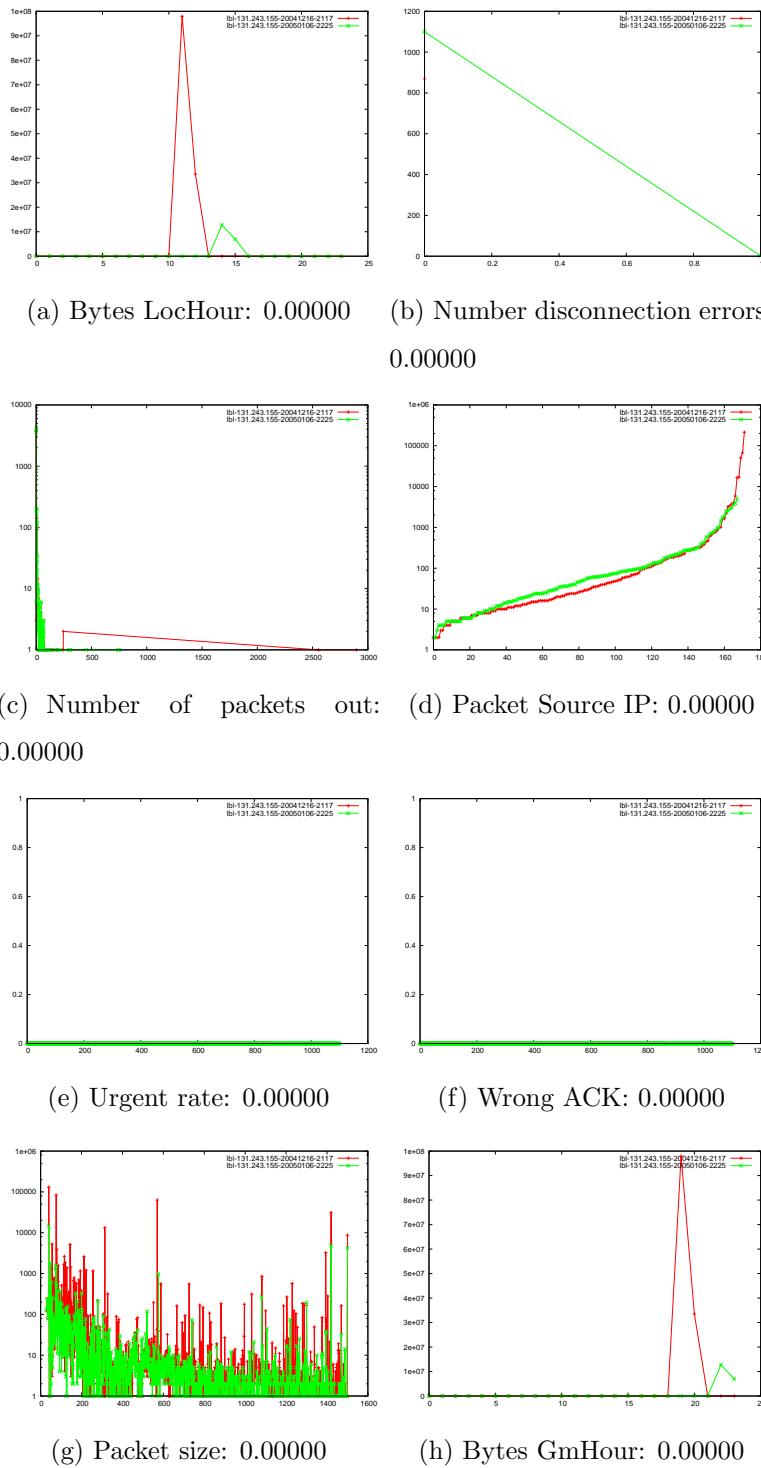


Figure M.7: Plots of characteristics 49 through 56 from basecase 3, pair 3

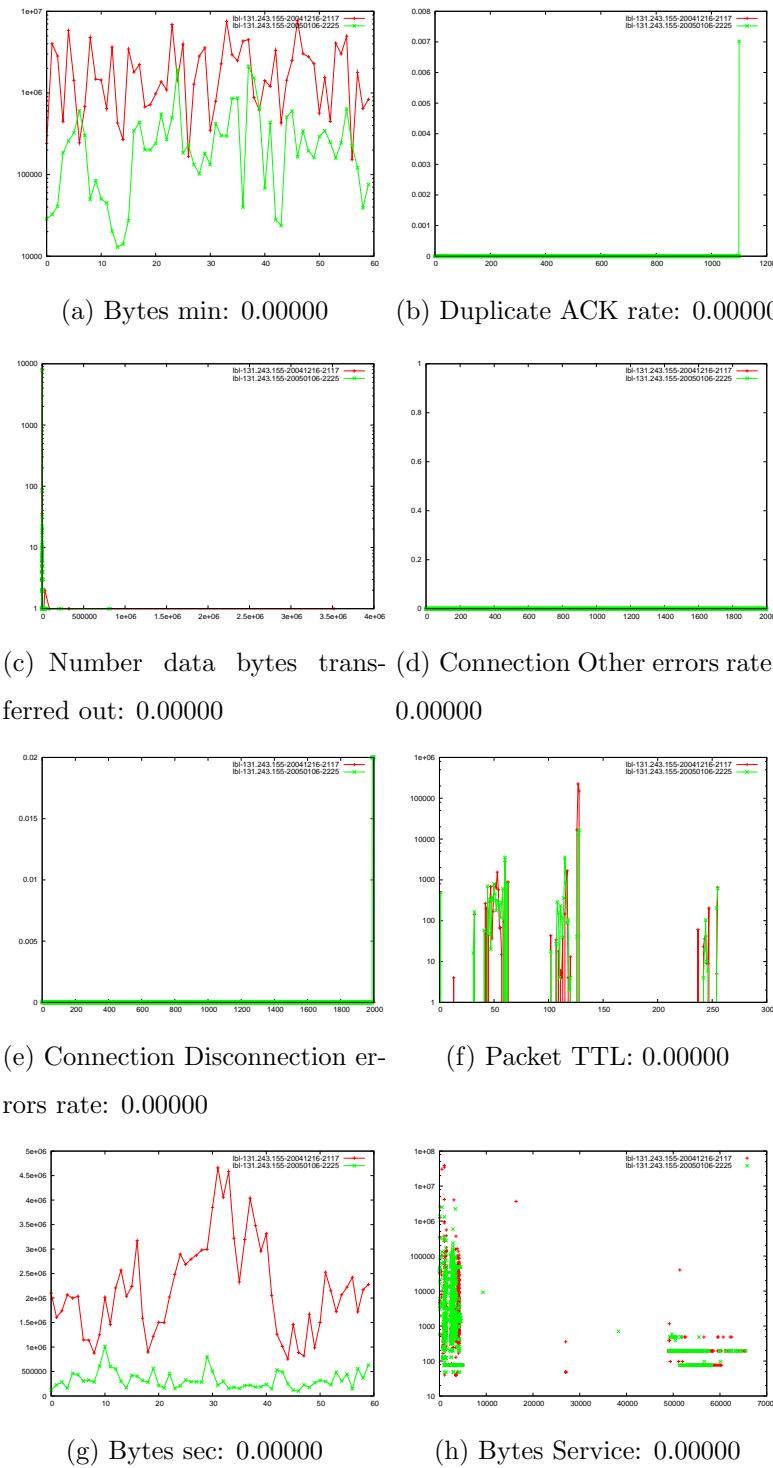


Figure M.8: Plots of characteristics 57 through 64 from basecase 3, pair 3

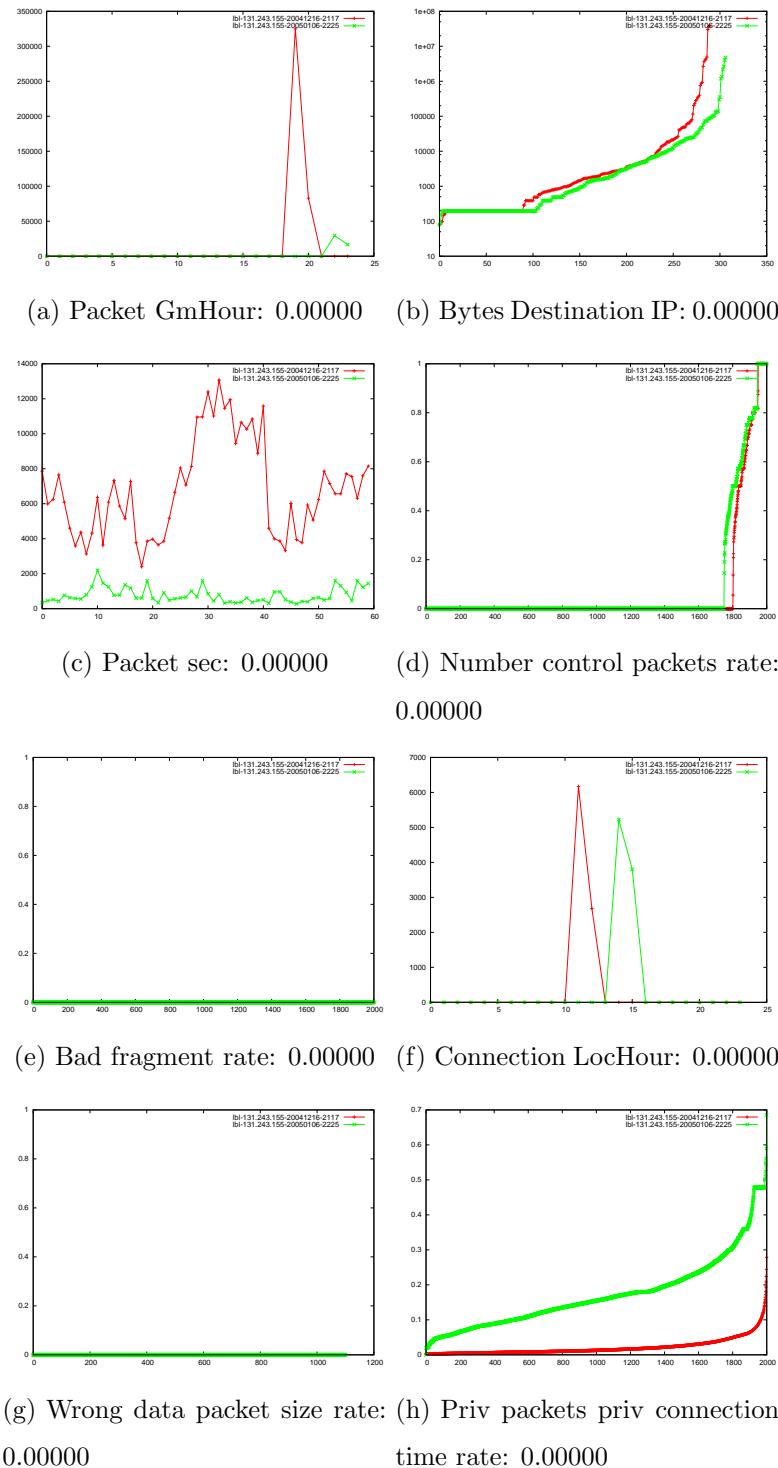


Figure M.9: Plots of characteristics 65 through 72 from basecase 3, pair 3

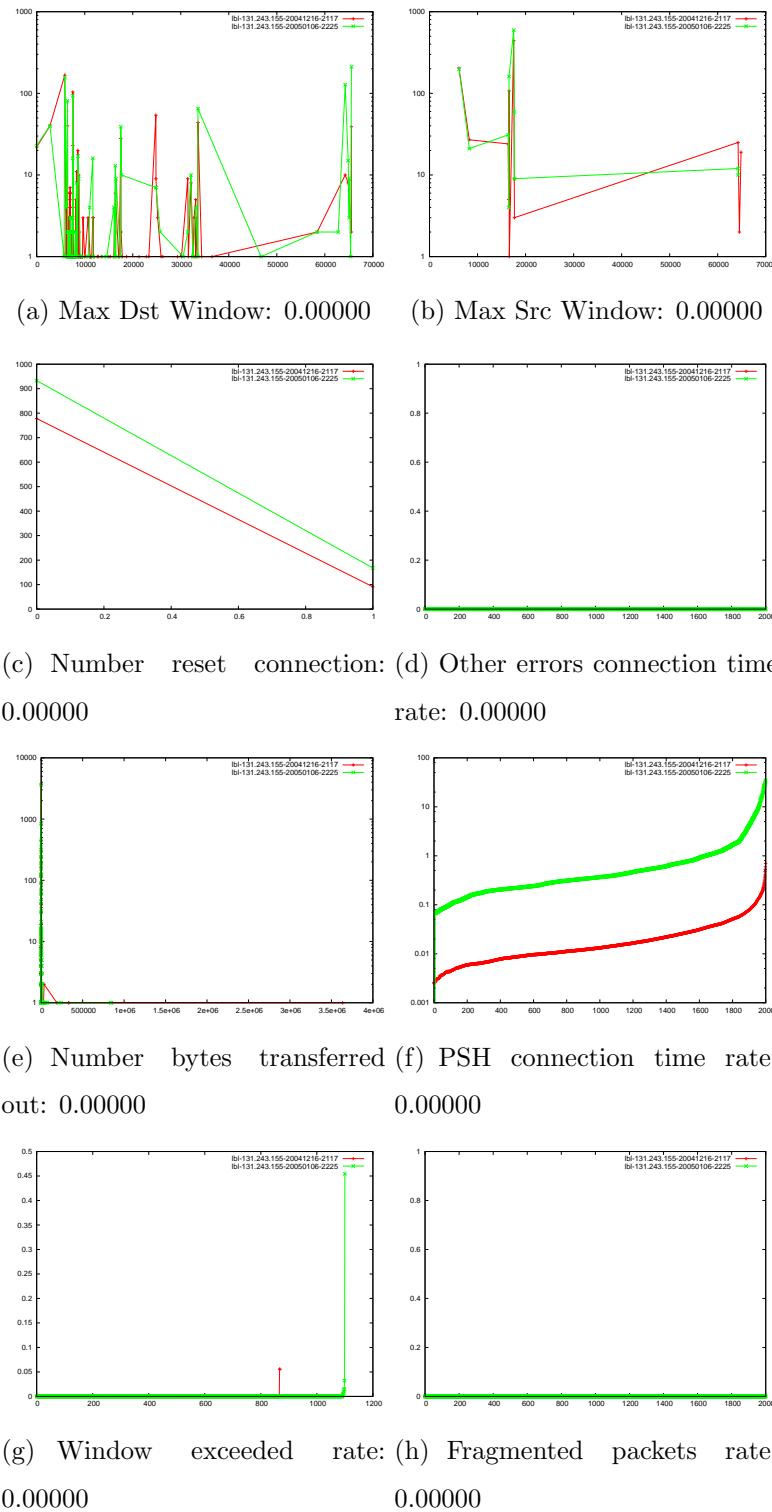


Figure M.10: Plots of characteristics 73 through 80 from basecase 3, pair 3

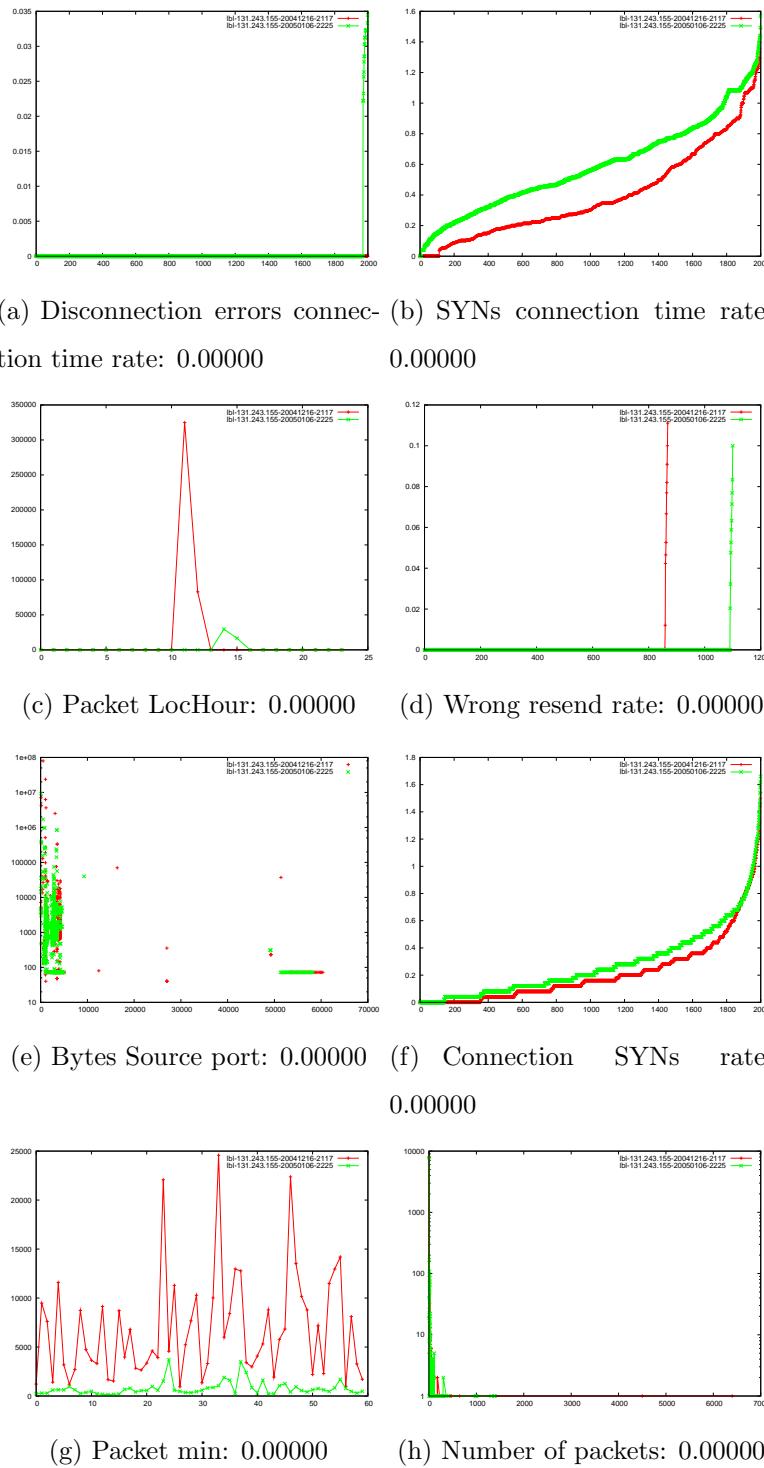


Figure M.11: Plots of characteristics 81 through 88 from basecase 3, pair 3

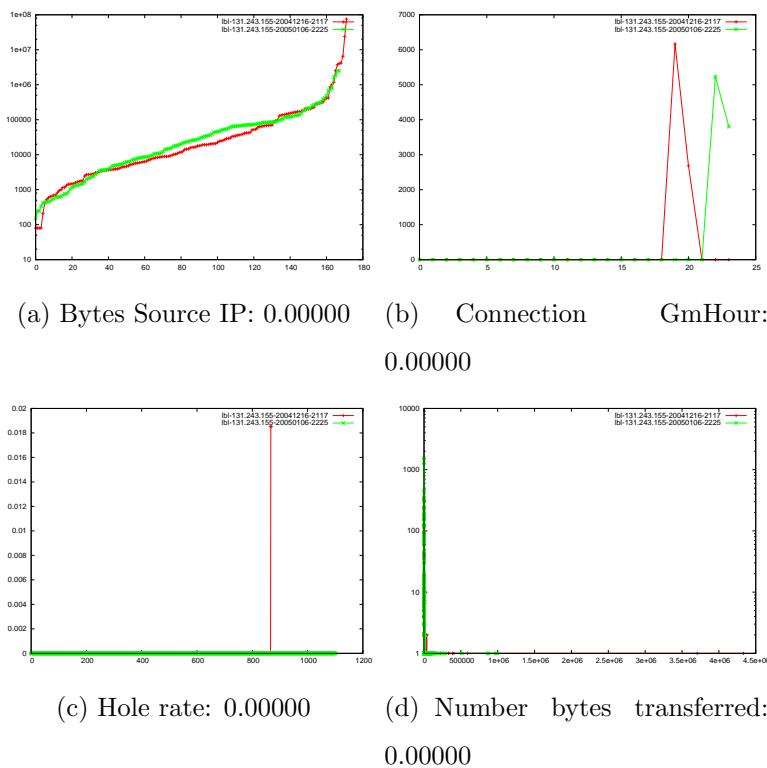


Figure M.12: Plots of characteristics 89 through 92 from basecase 3, pair 3

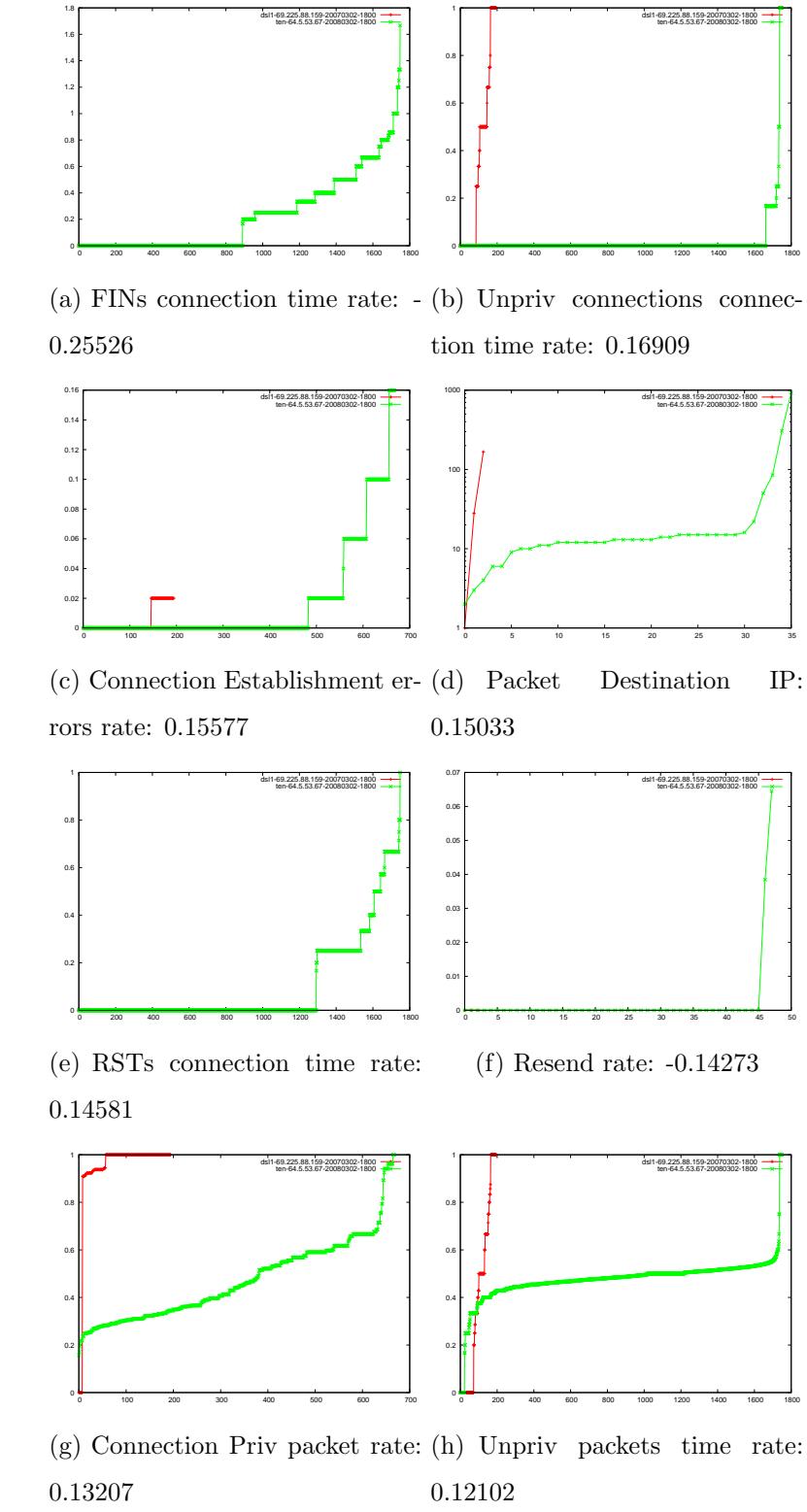
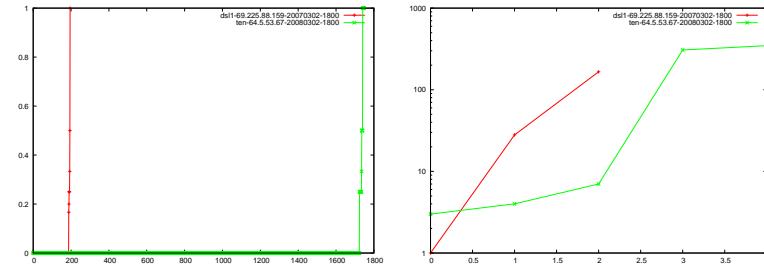
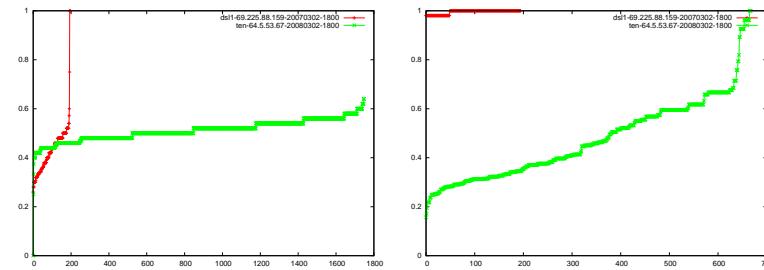


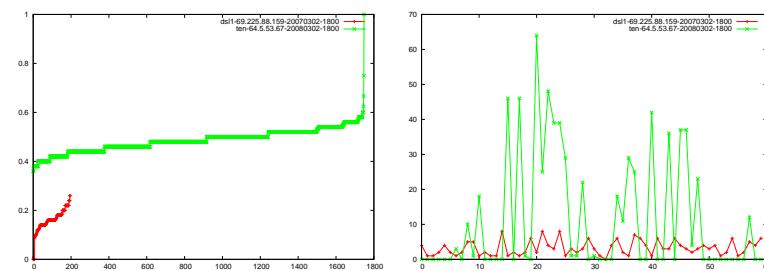
Figure M.13: Plots of characteristics 1 through 8 from basecase 8, pair 3



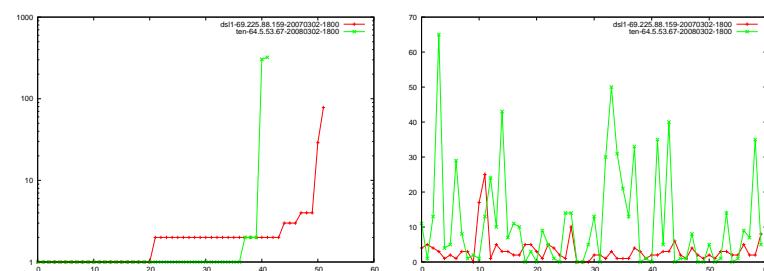
(a) Establishment errors connection time rate: -0.10114 (b) Connection Destination IP: - 0.09928



(c) Unpriv packets packet rate: 0.08201 (d) Connection packet rate: - 0.08096



(e) Priv packets packet rate: - 0.08085 (f) Connection min: -0.07491



(g) Connection Source IP: - 0.07442 (h) Connection sec: -0.07242

Figure M.14: Plots of characteristics 9 through 16 from basecase 8, pair 3

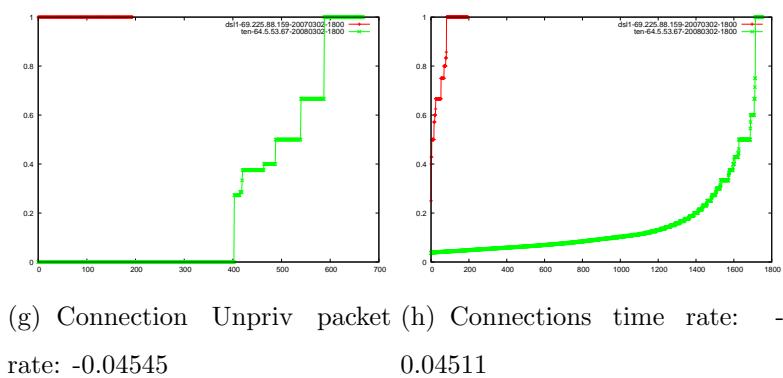
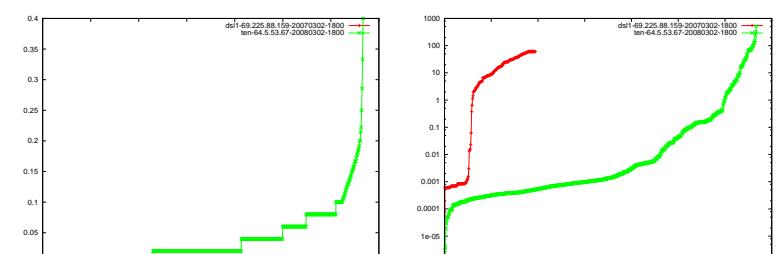
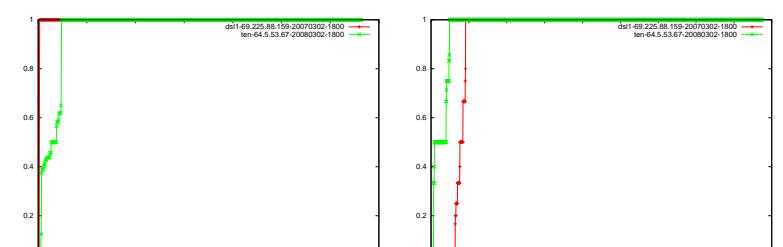
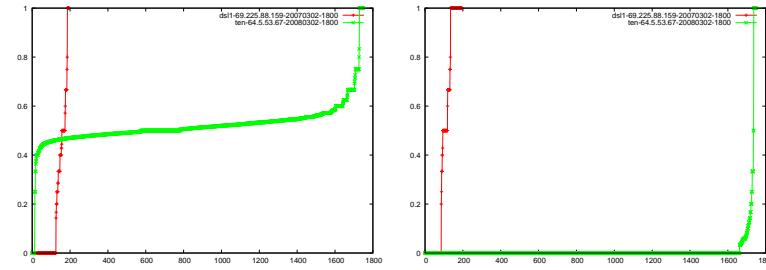
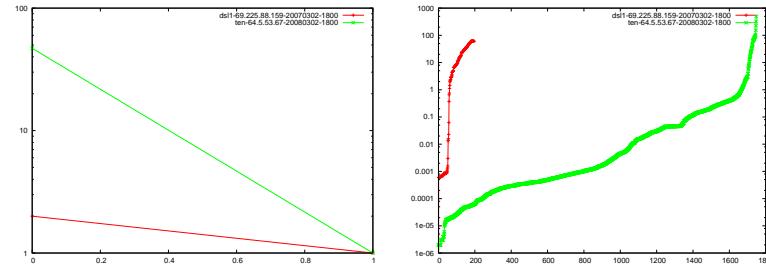
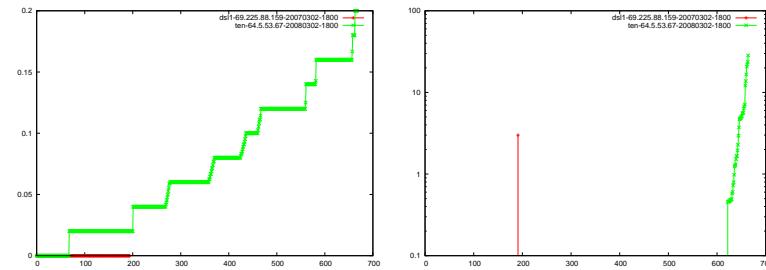


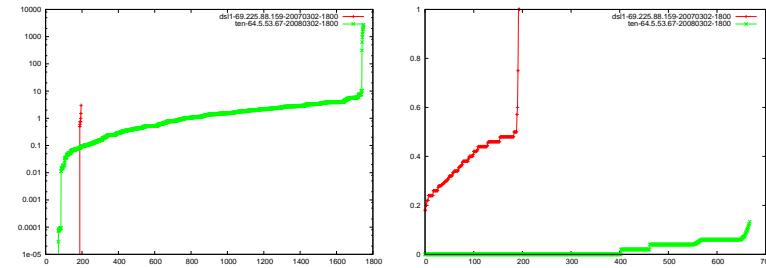
Figure M.15: Plots of characteristics 17 through 24 from basecase 8, pair 3



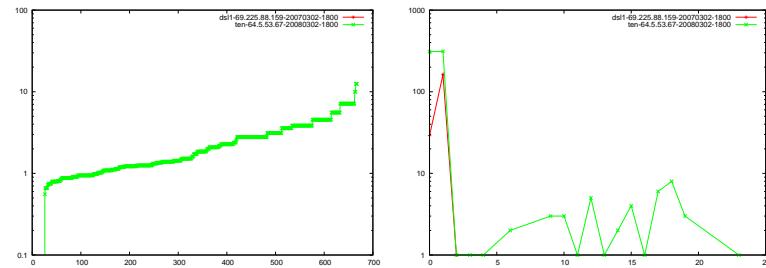
(a) Number connection errors: (b) InterPacket delta: -0.02447
0.03936



(c) Connection FINs rate: (d) Duration: 0.01622
0.02198



(e) Ave duration over last w secs: (f) Connection Unpriv connections rate: 0.00960
0.01133



(g) Connection PSH rate: (h) Number of packets in -
0.00888 0.00544

Figure M.16: Plots of characteristics 25 through 32 from basecase 8, pair 3

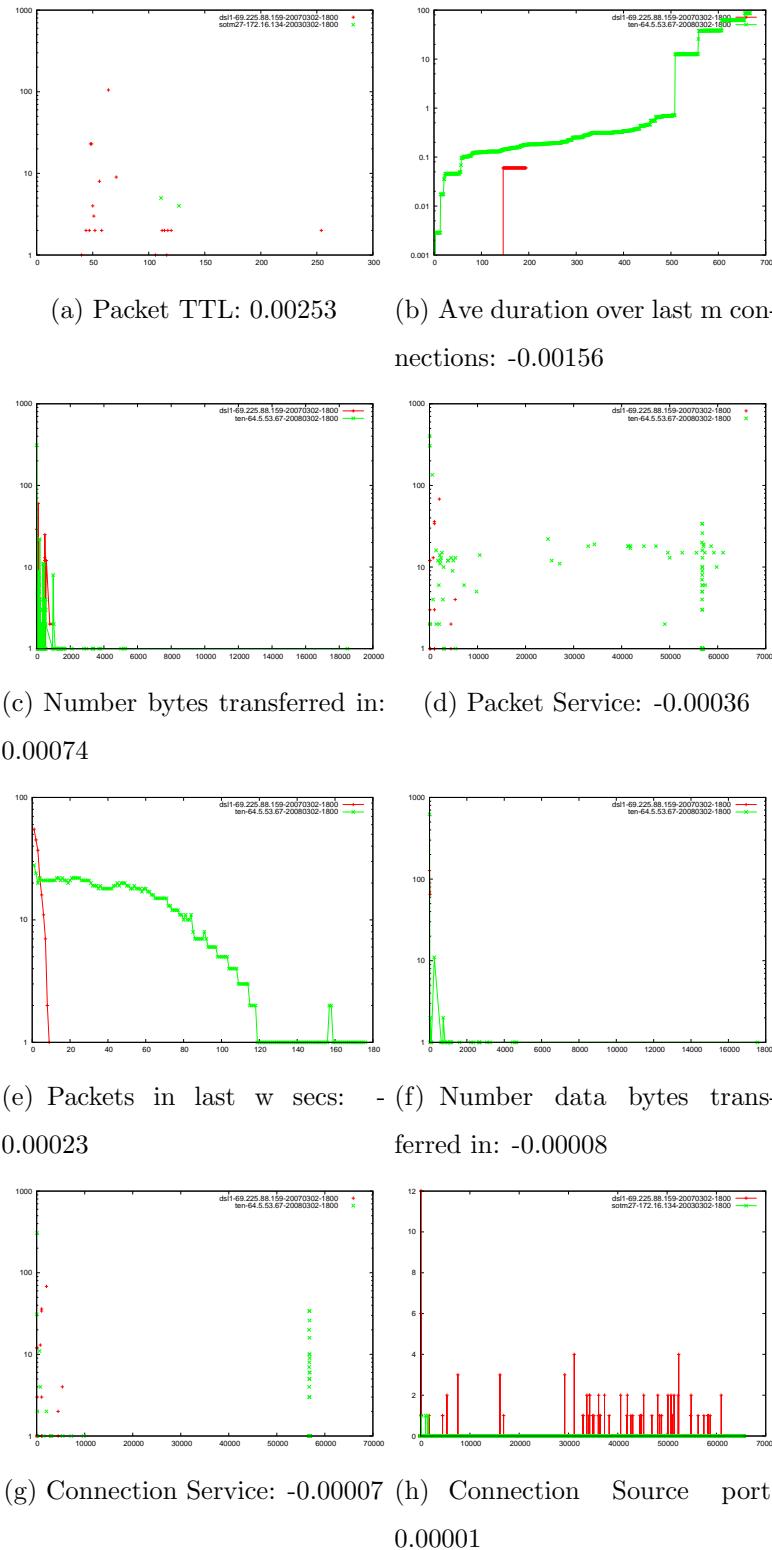
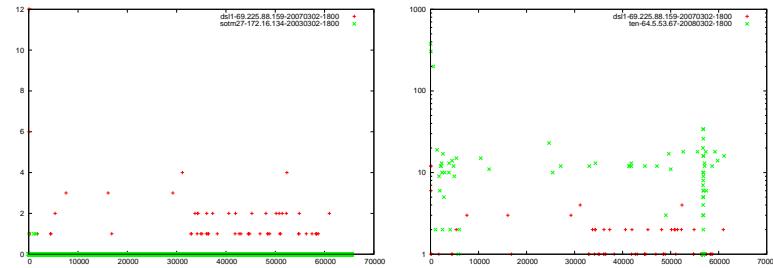
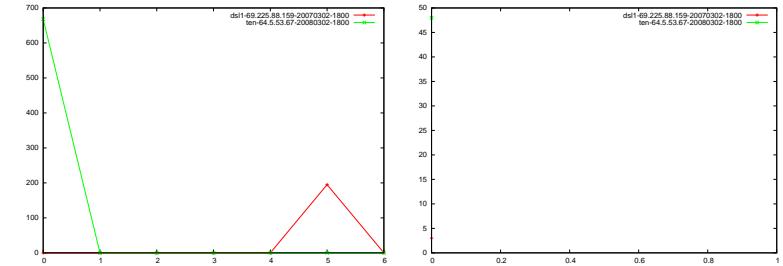


Figure M.17: Plots of characteristics 33 through 40 from basecase 8, pair 3

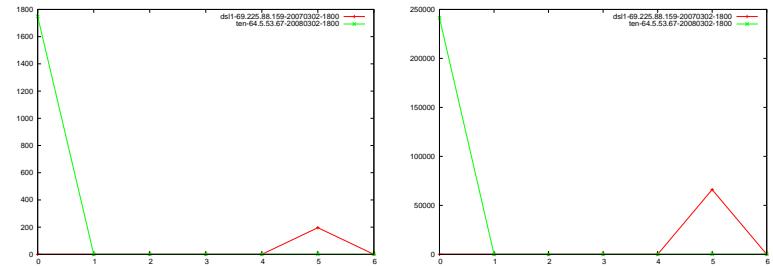


(a) Connection Source port: -
0.00001



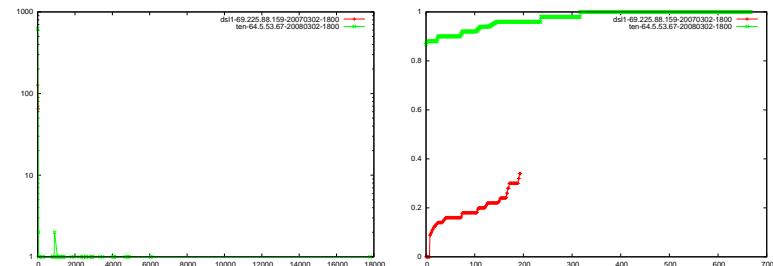
(c) Connection weekday: 0.00000

(d) Number other errors: 0.00000



(e) Packet weekday: 0.00000

(f) Bytes weekday: 0.00000



(g) Number data bytes transferred: 0.00000

(h) Connection Priv connections rate: 0.00000

Figure M.18: Plots of characteristics 41 through 48 from basecase 8, pair 3

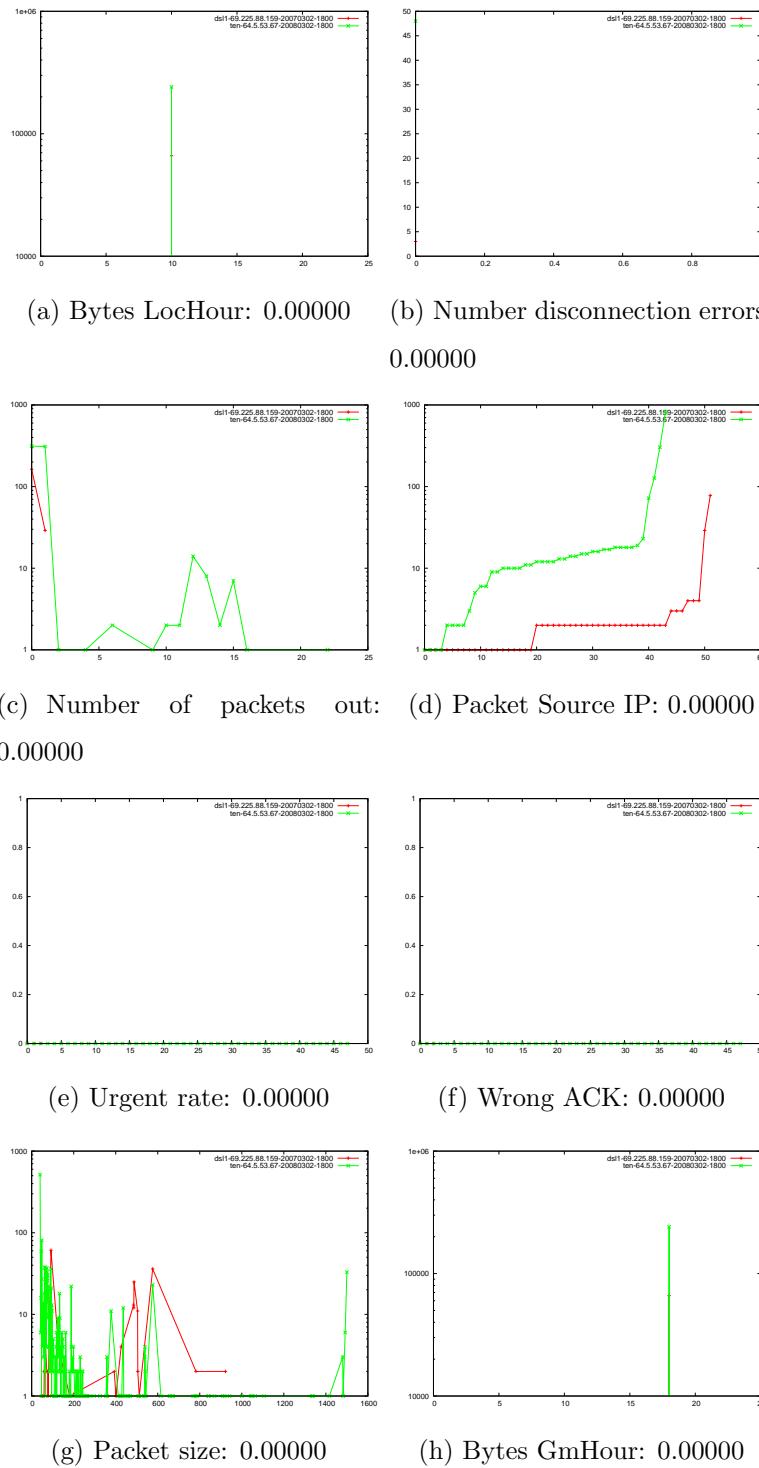


Figure M.19: Plots of characteristics 49 through 56 from basecase 8, pair 3

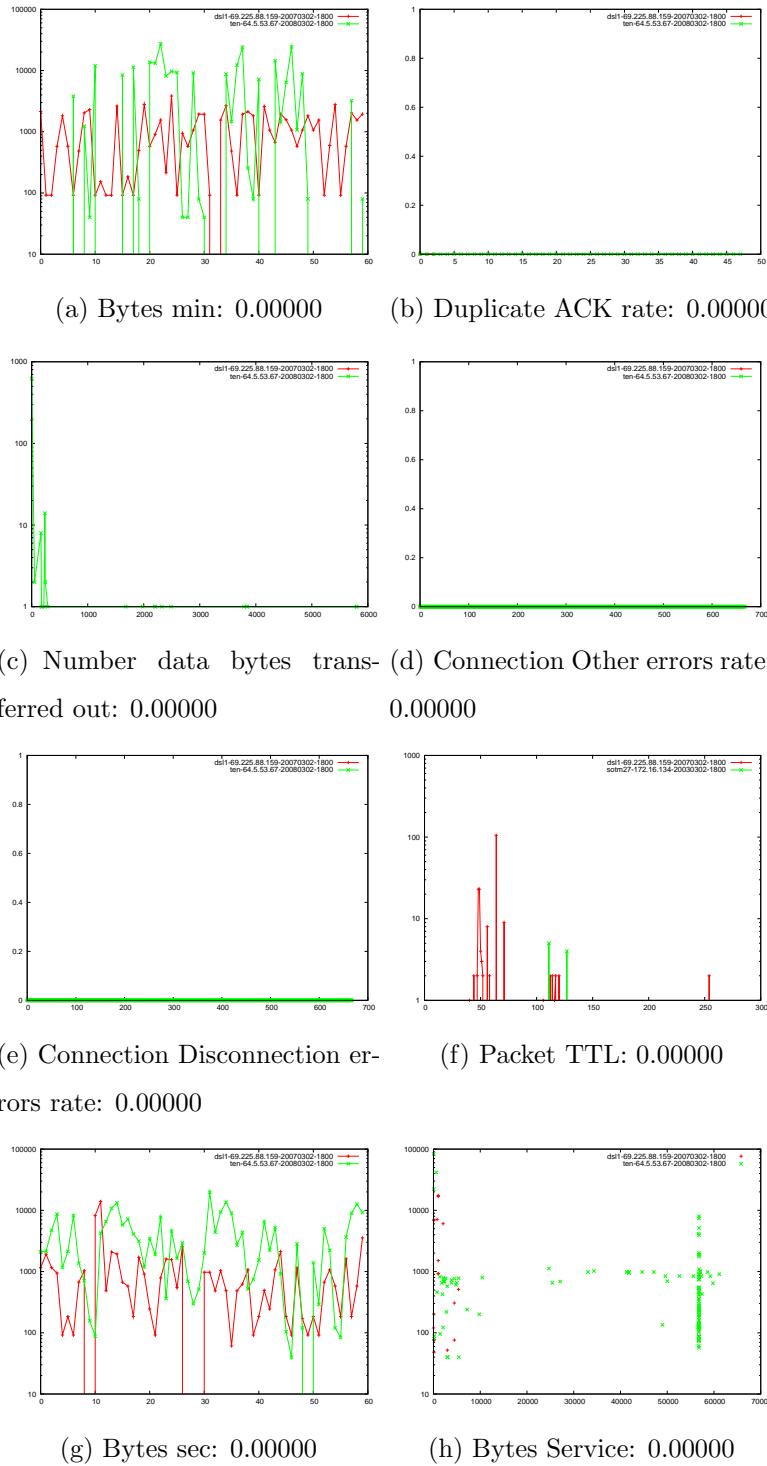


Figure M.20: Plots of characteristics 57 through 64 from basecase 8, pair 3

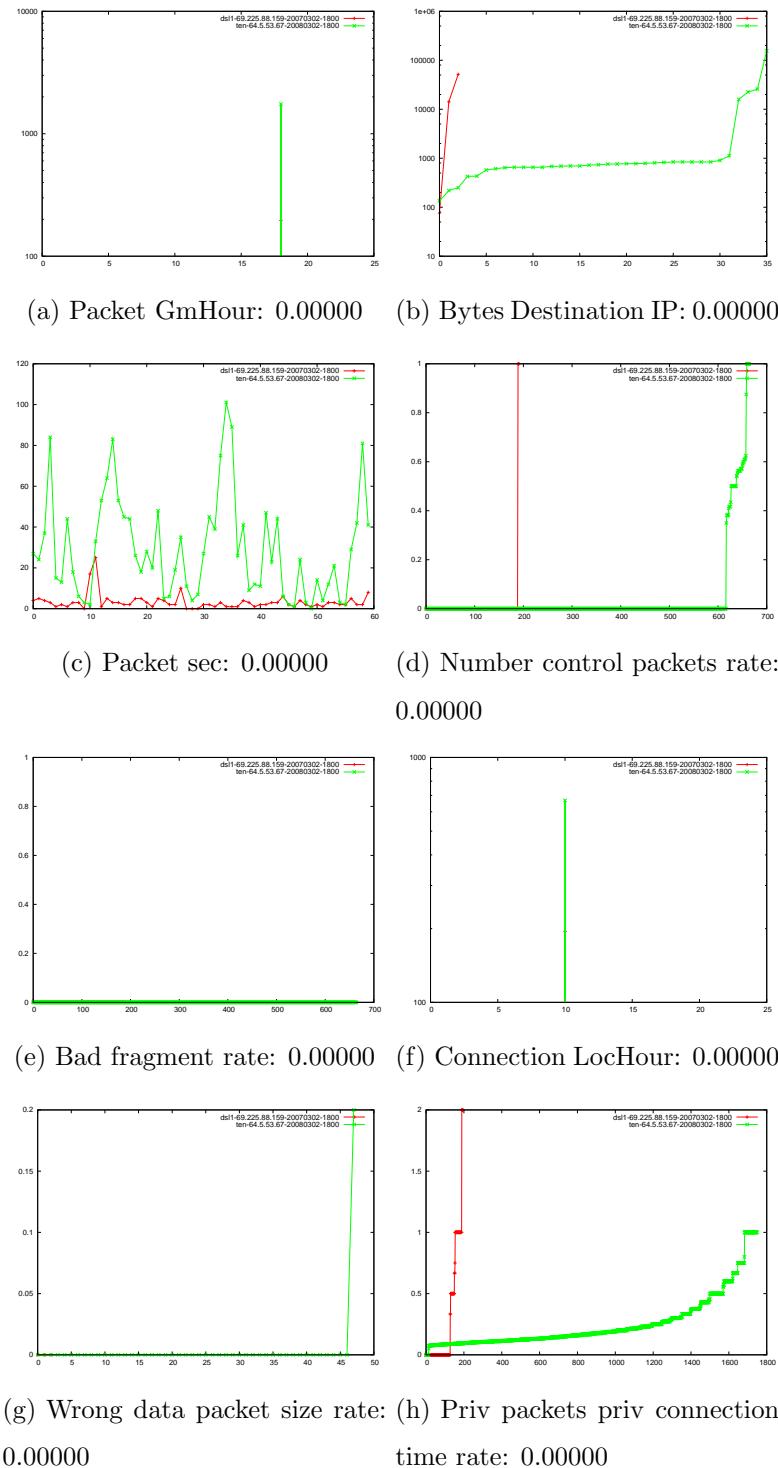


Figure M.21: Plots of characteristics 65 through 72 from basecase 8, pair 3

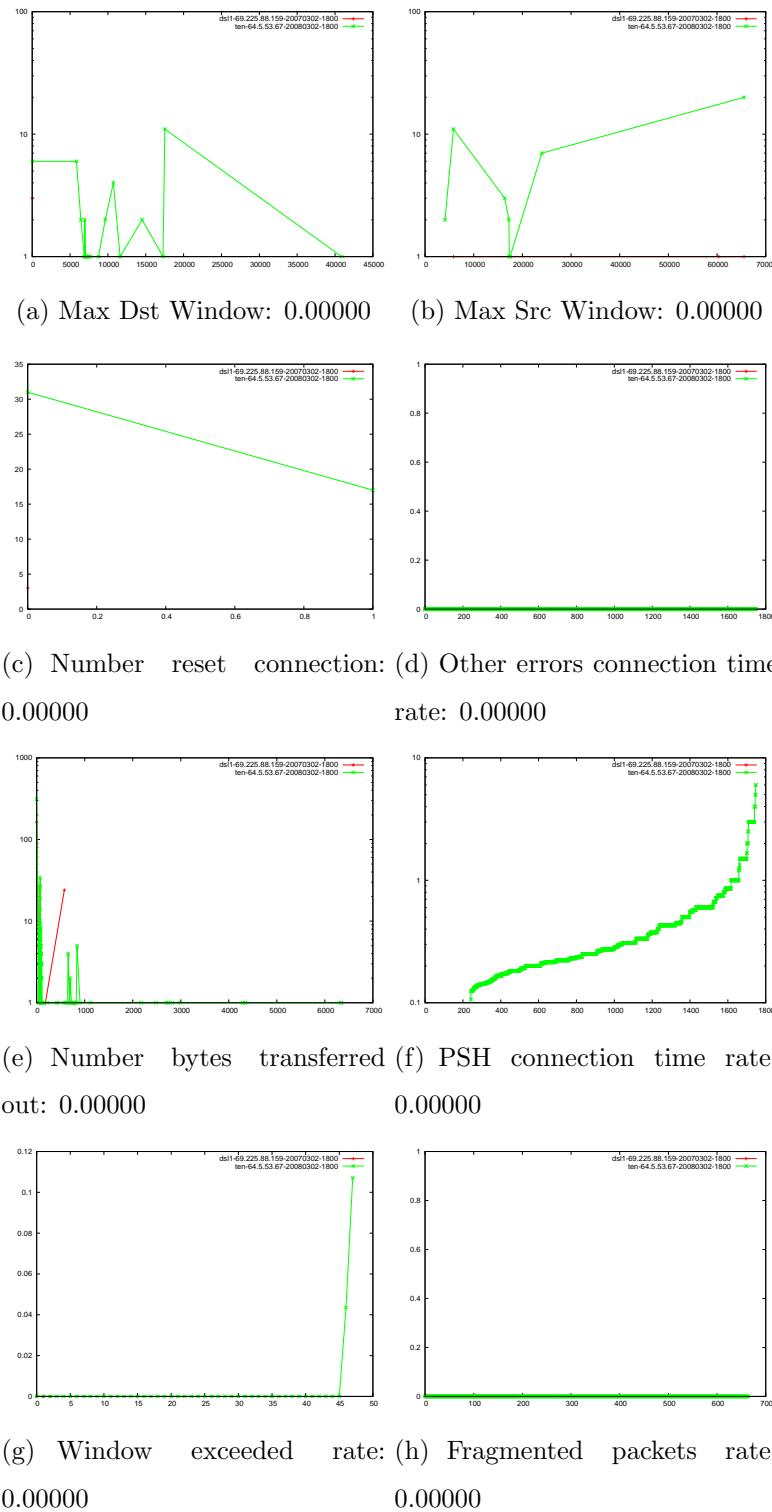


Figure M.22: Plots of characteristics 73 through 80 from basecase 8, pair 3

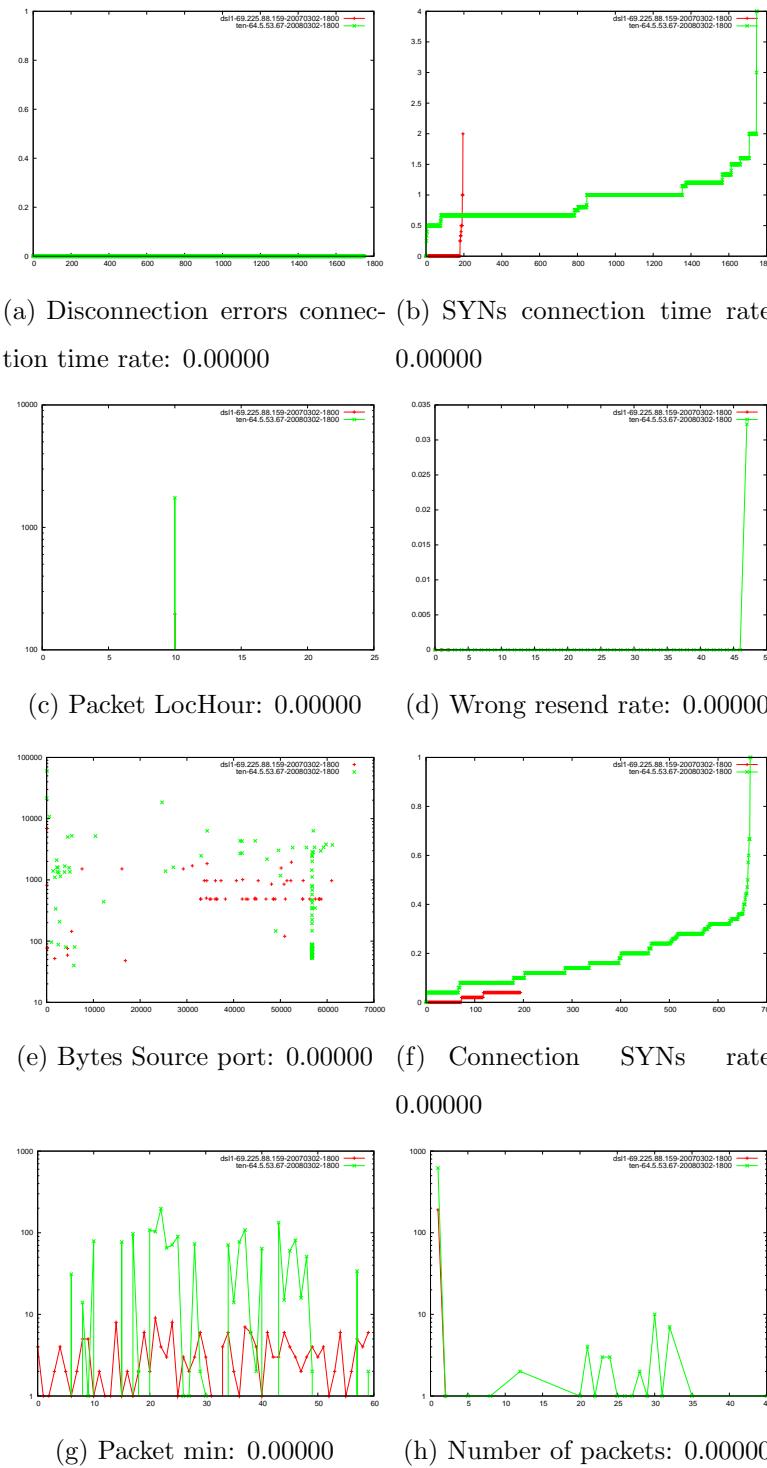


Figure M.23: Plots of characteristics 81 through 88 from basecase 8, pair 3

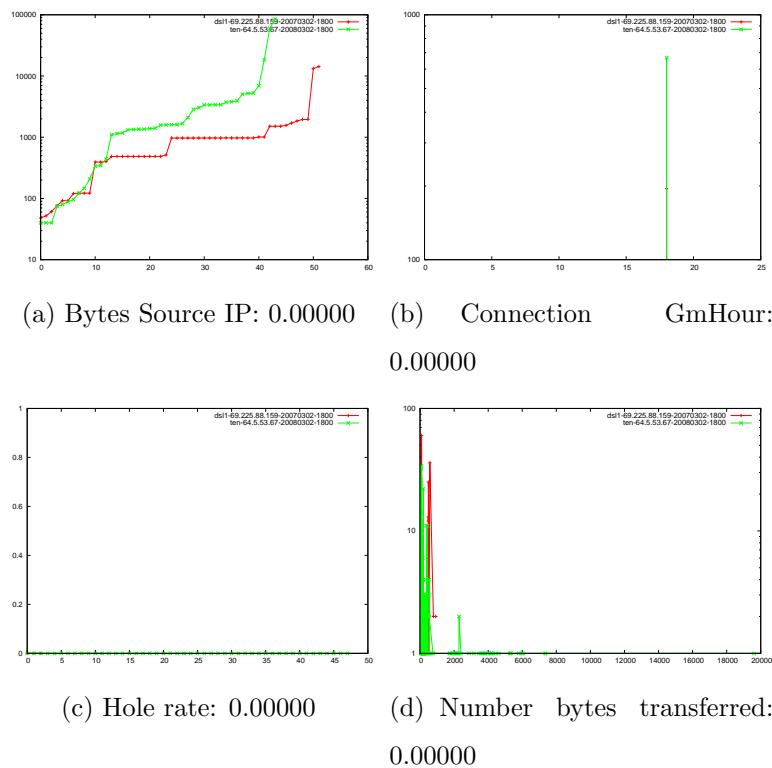


Figure M.24: Plots of characteristics 89 through 92 from basecase 8, pair 3