SANDIA REPORT
SAND2007-75757575
Unlimited Release
Printed November 2007

Characterizing and Improving Distributed Intrusion Detection Systems

Elliot P. Proebstel, Steven A. Hurd

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2007-75757575
Unlimited Release
Printed November 2007

Characterizing and Improving Distributed Intrusion Detection Systems

Elliot P. Proebstel
Department of Computer Science
The University of California, Davis
1 Shields Ave.
Davis, California 95616

Steven A. Hurd
Computer and Network Security Department
Sandia National Laboratories
Mail Stop (MS 9011)
P.O. Box 969
Livermore, California 94551-0969

Abstract

Due to ever-increasing quantities of information traversing networks, network administrators are developing greater reliance upon statistically sampled packet information as the source for their intrusion detection systems (IDS). Our research is aimed at understanding IDS performance when statistical packet sampling is used. Using the Snort IDS and a variety of data sets, we compared IDS results when an entire data set is used to the results when a statistically sampled subset of the data set is used. Generally speaking, IDS performance with statistically sampled information was shown to drop considerably even under fairly high sampling rates (such as 1:5).

Acknowledgements

The authors wish to extend our gratitude to Matt Bishop and Chen-Nee Chuah of UC Davis for their guidance and support on this work. Our thanks are also extended to Jianning Mai of UC Davis and Tao Ye of Sprint Advanced Technology Labs for their generous assistance.

We would also like to acknowledge our dataset sources, CRAWDAD and CAIDA, without which this work would not have been possible. Support for OC48 data collection is provided by DARPA, NSF, DHS, Cisco and CAIDA members.

Table of Contents

Table of Tables

1 Introduction

1.1 Background

Our work was originally aimed at understanding how well different intrusion detection systems (IDS) would work, using network traffic captured at different locations on a university campus network. This would allow us to identify which IDS work best with specific types of traffic (e.g. academic, residence halls, etc.). Furthermore, as the network traffic captured represents only a probabilistically-sampled fraction of the total network traffic, we would also be able to identify any shortcomings with traditional IDS when using probabilistically sampled network traffic.

This work was motivated by a desire to establish a baseline understanding of the accuracy issues likely faced by an enterprise network system administrator, many of whom face resource constraints that necessitate tradeoffs between capturing all packets (not technically feasible) and capturing a subset of packets (less desirable for security purposes). We intend to help inform those who need to make such tradeoffs in their network security design. Knowing the impact of these tradeoffs will allow network administrators to more intelligently apply their resources and interpret the output (and limitations) of their IDS.

1.2 Purpose

The overarching purpose of this investigation was to accurately characterize the challenges associated with conducting intrusion detection in a distributed environment.

Within that purpose, this investigation focused on issues relating to intrusion detection conducted using a statistically sampled subset of all network traffic. While traditional intrusion detection methods relied on analyzing all traffic, increasingly, statistical sampling of network traffic is being used in today's IDS. Thus, measuring IDS efficacy in a statistically sampled environment was the primary goal of this project.

To our knowledge, there have been no studies to date that establish the necessity of this advanced work for application in enterprise networks. That is, there are no published statistics on the accuracy degradation of current state of the art IDS when these systems are monitoring sampled data rather than full network traces. Furthermore, all known studies on this topic have been performed on backbone network traffic [1, 7, 8], the characteristics of which are often markedly different from those of an enterprise network, such as what might be found on a university campus.

1.3 Scope

The scope of this investigation is limited to comparing IDS efficacy under different rates of statistical sampling. Discussion in the "Approach" section (which follows this section) will articulate challenges and limitations to this line of inquiry.

—This page intentionally left blank —

2 Approach

2.1 Methods

The primary method of investigation was to compare results of IDS. We used multiple data sets, and within each data set, a range of statistical sampling rates, to best gauge performance degradation associated with statistical sampling.

Network Trace Information

We had initially planned upon using network trace information from a variety of sources. Most prominently, we planned to use information captured at a variety of points on the UC Davis campus network. Unfortunately, due to unforeseen administrative delays, we had little choice but to abandon this collection effort. Thus, inquiry was limited to use of pre-existing network trace information. The two sources, CRAWDAD and CAIDA, which represent enterprise and backbone traffic respectively, are described below.

For enterprise traffic, we are using network traces from the CRAWDAD (Community Resource for Archiving Wireless Data at Dartmouth) project at Dartmouth College [6]. These traces were gathered during the fall of 2003 through sniffing of the Dartmouth wireless traffic and thus represent only the traffic of wireless clients, which we consider a limitation of our dataset. However, the CRAWDAD datasets have provided us with an opportunity to compare IDS consistency on traffic observed in an academic building with traffic seen in residence halls. The traffic we are using was gathered between November 2, 2003 and February 28, 2004 in one academic hall and two residence halls (Residence Hall 100 and Residence Hall 13).

Our backbone traffic dataset is from CAIDA [3], the Cooperative Association for Internet Data Analysis. The traffic was observed on an OC48 link in San Jose. Though the traces we are using represent only a few hours of traffic on this link, they represent over 100 GB of raw network traffic.

As one might imagine, both CRAWDAD and CAIDA datasets are packet header traces that have been sanitized in a consistent (within the dataset), prefix-preserving manner. Thus, they are quite usable for the purposes of this inquiry.

Sampling Methodology

Each dataset is used in its original format as a "full trace"; what portscans an IDS detects in the full trace is considered a baseline for consistency on the trace. We then run the same IDS configuration on network traces in which the traffic has been sampled. We use a random packet sampling technique in which each packet is sampled with a probability of n and discarded with a probability of 1-n, where n is ranged to support sampling rates of 1:5, 1:10, 1:25, 1:50, 1:75, 1:100, 1:200, 1:500, and 1:1000. For any given trace, the sampled traces are

generated independently, meaning that a 1:10 sampled trace is not a strict subset of the corresponding 1:5 sampled trace.

Intrusion Detection System Information

In order to expand the relevancy of our research, as well as to provide baselines for comparison between our work and past research on anomaly detection in sampled traffic, we have performed an analysis of the consistency of the sfportscan preprocessor of SNORT [10], an open source IDS, on both enterprise and backbone traffic. We have run SNORT at low, medium, and high sense levels on all of our datasets. Aside from disabling detection for all other anomaly categories and changing the sfportscan sense level, we used default settings for the SNORT configuration.

Though we originally intended to compare SNORT's accuracy against at least one other IDS software package, we found our options were slim. Only two competing open source software packages even claimed to perform portscan detection on pcap-formatted network traces. The first, Firestorm [5], has not been updated since 2004. It proved impossible to install, presumably because it has not kept pace with updated system libraries.

We are still in the process of trying to get meaningful output from the other competing product, Bro [2]. One significant hindrance to this effort was the discovery that Bro does not support the link layer protocol Cisco HDLC (cHDLC), which was used in the capture of the CAIDA traces. Additionally, installation and configuration of Bro have proven significantly more challenging than were installation and configuration of SNORT. The extensive and detailed documentation for SNORT made it a much easier product to manage and use.

Determining Consistency Rates

To establish the *consistency rates* of an IDS on a particular dataset, we compare alerts raised by that IDS on the full trace with alerts raised by that IDS on the associated sampled traces. Every alert from a sampled trace is compared with alerts from the full trace. Matches are generously assumed to be present when an alert from the sampled trace has the same relevant IP address, general scan type, and time window (60 seconds for the low sense level, 90 for the medium sense level, and 600 for the high sense level). The full code used to detect matches between sampled trace alerts and full trace alerts is provided in Appendix E of this report.

2.2 Assumptions

A critical assumption in this research is that short of manually analyzing each network trace in full, there is no definitive method of establishing "ground truth" regarding an IDS' accuracy in detecting portscans.

For example, there may be scans that SNORT fails to detect in the full network traces – false negatives – and there may likewise be alerts raised by SNORT that relate to legitimate non-scanning network traffic – false positives. However, SNORT's wide use in the security community suggests that its accuracy is acceptable, and furthermore, our intention is not

necessarily to establish the baseline accuracy of an IDS; rather, we concern ourselves with the question of whether an IDS will perform consistently on full traces and their corresponding sampled traces. Thus, for this study we have operated under the assumption that a scan detected by SNORT in the full trace is a *real scan* and that an alert raised by SNORT in a sampled trace that does not relate to a *real scan* is a false positive, while a *real scan* that has no corresponding alerts from the sampled traces is recorded as a false negative for that sampled trace.

In order to disambiguate true false positives and false negatives (that is, those which relate the accuracy of SNORT to ground truth) from the inconsistencies arising between SNORT's alerts on sampled traces and the *real scans*, we use the term *false scan* to reference an alert raised on a sampled trace that does not relate to a *real scan* and the term *missed scan* to reference a *real scan* that has no corresponding alerts in the relevant sampled trace. As a simple case, assume that SNORT detected 5 portscans in a particular full trace and then raised only 2 portscan alerts on the 1:5 sampled trace. We compare the alerts from the sampled trace to the 5 real scans and find that one alert relates to a *real scan*, while the other cannot be correlated to any of the *real scans*. We say here that the sampled trace had 1 *false scan* and 4 *missed scans*, yielding a *missed scan rate* of 80%, a *consistency rate* of 20%, and a *false scan rate* of 50%.

2.3 Procedures

- Prepare datasets
  1. Acquire permission from CAIDA and CRAWDAD to use datasets
  2. Download and organize datasets into relevant categories
  3. Perform random sampling of datasets to generate sampled traces

- Prepare IDS
  1. Download and install SNORT
  2. Configure SNORT for testing
     a. Disable unrelated detection schemes

- Perform IDS analysis of datasets
  1. Configure SNORT's sfportscan sense level to low
  2. Run SNORT on full traces and sampled traces
  3. Repeat 1-2 for sense levels of medium and high

- Determine consistency of IDS
  1. Compare alerts from sampled traces to alerts from full traces
  2. Analyze results

—This page intentionally left blank —

3 Results and Discussion

We have found that SNORT's *consistency rates* differ significantly between sense levels and between datasets. In our analysis, we discuss *scans* generically, including both portscans and portsweeps. It was often the case that a dataset contained a statistically insignificant number of scans of one type but a very large number of scans of the other type. Appendix D of this report presents full tables documenting the exact observations from our experiments, while this discussion will report the broad lessons gained from our research and supply a few abbreviated tables for support.

At the low sense level, detection rates drop so sharply that even at a sampling rate of 1:5, the highest *consistency rate* (observed in the academic network traffic) was just below 12%, while the other network traces were between 1% and 8%. The accompanying *false scan* rate at the 1:5 sampling rate ranged from 4% to over 77%. With a sampling rate of 1:10, all *consistency rates* fell below 3% with *false scan rates* ranging between 5% and 100%.

Table 3.1 Excerpt of CRAWDAD Academic Hall Portsweeps at low sense level.

Sampling rate	# of scan alerts	# of *false scans*	*False scan rate*	*Detected scans*	# of *missed scans*	*Missed scan rate*	*Consistency rate*
1:1	8221	0	0.00%	8221	0	0.00%	100.00%
1:5	1207	234	19.39%	973	7248	88.16%	11.84%
1:10	235	39	16.60%	196	8025	97.62%	2.38%
1:25	6	0	N/A	6	8215	99.93%	0.07%

On the medium sense level, detection rates for the backbone traces still dropped sharply to 12% at a 1:5 sampling rate, but *false scan rates* stayed below 5% for all sampling rates, demonstrating a marked improvement over the *false scan rates* we observed in the low sense level. Even more encouraging, the *consistency rate* for the academic network traffic was over 80% at the 1:5 sampling rate and even stayed above 70% for sampling rates of 1:10 and 1:25 – while displaying very low *false scan rates*.

Table 3.2 Excerpt of CRAWDAD Academic Hall Portsweeps at medium sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	35911	0	0.00%	35911	0	0.00%	100.00%
1:5	32621	2965	8.26%	29656	6255	17.42%	82.58%
1:10	27533	367	1.33%	27166	8745	24.35%	75.65%
1:25	25624	38	0.15%	25586	10325	28.75%	71.25%
1:50	19117	11	0.06%	19106	16805	46.80%	53.20%

In our testing of the high sense level, we have demonstrated again a sharp drop in *consistency rates* for the backbone traffic and the residence hall traffic – both falling to approximately 10% with a 1:5 sampling rate, dipping below 5% with 1:10 sampling, and finally disappearing below 2% for all remaining sampling rates. We did, however, again observe low *false scan rates* for all sampling ranges; these rates stayed in the single digits across all sampling rates. For the academic network traffic, the high sense level displayed the highest resistance to sampling impact. *Consistency rates* were nearly 60% at 1:5, over 30% at 1:10, and stayed in the 15% range for sampling rates ranging from 1:25 to 1:200. *False scan rates* stayed very low for all these ranges, as well.

Table 3.3 Excerpt of CRAWDAD Academic Hall Portsweeps at high sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	31480	0	0.00%	31480	0	0.00%	100.00%
1:5	19124	704	3.68%	18420	13060	41.49%	58.51%
1:10	10458	252	2.41%	10206	21274	67.58%	32.42%
1:25	6194	32	0.52%	6162	25318	80.43%	19.57%
1:50	5333	10	0.19%	5323	26157	83.09%	16.91%

SNORT configuration instructions recommend that medium and high sense levels be used only with manual sfportscan tuning, as the higher sense levels often generate more false positives [11]. In all cases, we used the default settings for the various sense levels. SNORT's sfportscan low sense level detects scans based only on the number of RST (connection reset) responses a host receives in a given time window [9]. Given that an active benign host is unlikely to contact a large number of unavailable hosts or services, the low sense level is particularly unlikely to generate many true false positives. However, this also makes the low sense level very sensitive to the effects of random packet sampling, which is known to present a strong bias in favor of longer flows and miss a large percentage of short flows [4]. Obviously, a flow consisting only of a SYN (connection request) packet and an RST packet is very short and would easily be missed by random packet sampling, which helps explain why the low sense level experienced such a high percentage of missed scans.

Medium and high sense levels not only expand the time windows for enumerating RST responses, but they also make use of connection counts per host [9]. By tracking connection counts, these sense levels can detect scans launched against firewalled hosts, but they are also much more likely to raise false alarms on highly active benign hosts. For this reason, the SNORT Reference Manual recommends that an operator review alerts for such false positives and reactively update the sfportscan configuration to ignore such active hosts in the future to avoid clogging alert logs [11].

As discussed previously, we do not have the capacity to manually verify whether the alerts raised on the full traces are true positives in the sense of ground truth, as the data available to us is only sanitized packet headers. However, our experiments have demonstrated that SNORT does not scale well on sampled data. Alerts raised on sampled data are not a representative sample of the alerts that would be raised on the full network trace from which those samples were extracted. Not only do *consistency rates* suffer tremendously, but the number of *false scans* tends to be very high.

We have also shown SNORT to be most resistant to the impact of sampling when applied against traffic traversing an academic network using a medium sense level and a sampling rate of at least 1:100. SNORT is also capable of performing moderately well on academic traffic using a high sense level and a sampling rate of at least 1:200. SNORT's consistency, at all sampling rates and with all sense levels, degraded rapidly on both backbone traffic and residence hall traffic.

—This page intentionally left blank —

4 Conclusions

Simply put, the results strongly indicate that the use of statistical sampling in IDS usage causes IDS consistency to deteriorate significantly.

Certainly, there are cases where statistically sampled IDS performance was reasonably good. The example of a CRAWDAD Academic Hall data, at 1:5 sampling rate and medium sense level, produced fairly accurate alerts (82.58% of the original alerts). On the other hand, that same 1:5 sampling rate used with low or high sense levels yielded far less satisfactory results (11.84% and 58.51% respectively). It is notable that a 1:5 sampling rate is considerably less than the 1:1000 sampling which appears to be commonly used.

From these results, it is clear that IDS performance is very sensitive to tuning issues. While an IDS analyst may not be able to dictate the sampling rates, they can adapt other parameters (such as the sense level) to match the needs of their environment.

It should be noted that due to the amount of malicious traffic on the internet, relatively benign malicious traffic (such as simple port scans) are not likely to attract significant attention from an IDS analyst. Thus, it may be reasonable to question the applicability of these results.

That said, detecting a port scan may be the simplest thing an IDS is called upon to do. While it often requires keeping state information (rather than merely checking a packet against signatures), it represents the simplest in multi-packet IDS actions. For other attacks that are only detected as a result of stateful analysis, it follows that statistical sampling greatly reduces the performance of an IDS.

As mentioned in section 2.1, it is regrettable that we were unable to secure UC Davis campus information for this analysis. However, even had we used that information for the analysis, it seems unlikely that the results and conclusion would vary significantly from those presented.

Finally, as mentioned earlier, this research is based upon the assumption that the 1:1 sampling rate detections represent “ground truth”. While it seems intuitive that 1:1 sampling provides the highest fidelity results, it is possible that some or all of the alerts generated through 1:1 sampling, but not generated at higher sampling rates, represent false positives (in terms of ground truth). This is a thorny issue and is the core problem of IDS. IDS analysts are continuously forced to discard false positive (or less relevant true positives) in the course of their work. It is only through retrospective analysis in conjunction with an experienced IDS analyst can we determine what information generated by an IDS was the most critical information. Furthermore, while false positives are a significant issue with which an IDS analyst must deal, false negatives can be an even greater risk to an institution – especially if the analyst does not have a way of knowing how many false negatives are occurring. Our work is a step in the direction of providing guidance to the analyst so that organizations which are constrained into using sampled data for IDS purposes at least have a sense of how many anomalies their IDS may be missing.

—This page intentionally left blank —

5 Recommendations

As the need for use of statistical sampling techniques in IDS is likely to continue to increase, further research in this area would seem to be indicated.

One new direction would be to take data sets including actual intrusions and run them through a series of IDS using a variety of sampling rates. As identifying actual intrusions would likely be a more critical task for IDS than merely identifying a port scan, this would provide more relevant feedback as to the criticality of the sampling issue. The work in [1] is similar but does not directly address the need to assess existing IDS software for consistency.

Along those lines, it would be instructive to have an experienced IDS analyst that is familiar with the "ground truth" of the underlying dataset examine the IDS results (given different sampling rates) and identify detections and missed detections, as well as false positives.

Given the difficulty we faced in securing datasets of sanitized packet headers only, we imagine that this call to action for future research may need to be answered by industry. An ideal future direction for this work would be to use full network traces with a current IDS installation. This would answer the issues of preprocessor tuning, configuration settings, and operator unfamiliarity that we faced.

We hope that our work provides motivation for further research into IDS software that can support sampled network traces. The promising work being done by UC Davis and Sprint Advanced Technology Labs [7, 8] suggests that there are directions for this work that could lead to exciting new IDS developments. However, as our work demonstrates, solutions which work well for enterprise network traffic do not always work as well for backbone traffic – and the reverse is likely true. Thus, industry would do well to embark on similar, but possibly divergent, paths to ensure that IDS software can keep pace as enterprise network use continues to grow.

—This page intentionally left blank —

Appendix A: References

[1] Braukhoff, D., Tellenbach, B., Wagner, A., May, M., and Lakhina, A. Impact of packet sampling on anomaly detection metrics. In *Proceedings of the ACM Internet Measurement Conference* (October 2006), pp. 159-164.

[2] Bro Intrusion Detection System. http://www.bro-ids.org/.

[3] CAIDA data set OC48 Link A (San Jose, CA). Downloaded from http://www.caida.org/data, Feb 2007.

[4] Duffield, N., Lund, C., and Thorup, M. Estimating flow distributions from sampled flow statistics. In *Proc. ACM/SIGCOMM*, 2003, pp. 325-336.

[5] Firestorm. http://sourceforge.net/projects/firestorm-ids.

[6] Kotz, D., Henderson, T., and Abyzov, I. CRAWDAD data set dartmouth/campus/tcpdump/fall0304 (v. 2004-11-09). Downloaded from http://crawdad.cs.dartmouth.edu/dartmouth/campus/tcpdump/, Feb 2007.

[7] Mai, J., Chuah, C.-N., Sridharan, A., Ye, T., and Zang, H. Is sampled data sufficient for anomaly detection ? *IMC 2006* (Rio de Janeiro, Brazil, October 2006).

[8] Mai, J., Sridharan, A., Chuah, C.-N., Zang, H., and Ye, T. Impact of packet sampling on portscan detection. *IEEE Journal on Selected Areas in Communication* (2006).

[9] Malmedal, Bjarte. Using netflows for slow portscan detection. Master's theis, Department of Computer Science and Media Technology, Gjovik University College, 2005.

[10] SNORT. http://snort.org.

[11] SNORT Reference Manual. Accessed at http://www.snort.org/docs/snort_htmanuals/htmanual_2.4/node11.html. Nov 2007.

—This page intentionally left blank —

Appendix B: Acronyms, Symbols, Abbreviations

CAIDA	Cooperative Association for Internet Data Analysis
CRAWDAD	Community Resource for Archiving Wireless Data at Dartmouth
HDLC	High-Level Data Link Control
IDS	Intrusion Detection System

—This page intentionally left blank —

Appendix C: Glossary

Consistency Rate	We define this term as the percentage of *real scans* that are detected by the IDS on a given sampled trace. It is calculated by dividing the number of *detected scans* by the number of *real scans* for the particular trace. It is also inversely proportionate to the *missed scan rate.*
Detected Scan	We define this as a *real scan* for which an alert is raised by an IDS on a particular sampled trace.
False Negative	An instance of anomalous traffic for which an IDS fails to raise an alert.
False Positive	An alert raised by an IDS that does not relate to anomalous traffic but, rather, is a result of benign traffic.
False Scan	We define this as an alert raised by an IDS on a sampled trace that does not correlate to a *real scan*.
False Scan Rate	We define this term as the percentage of alerts raised by an IDS on a sampled trace that are *false scans*. It is calculated by dividing the number of *false scans* by the number of *real scans* for the particular trace.
Missed Scan	We define this as a *real scan* for which no alert is raised by an IDS on a particular sampled trace.
Missed Scan Rate	We define this term as the percentage of *real scans* that are missed by the IDS on a given sampled trace. It is calculated by dividing the number of *missed scans* by the number of *real scans* for the particular trace. It is also inversely proportionate to the *consistency rate*.
Real Scan	We define this as a scan that is observed by the IDS in a full network trace.

—This page intentionally left blank —

Appendix D: Full SNORT Results

This appendix provides the full results of SNORT sfportscan detection on the CAIDA and CRAWDAD datasets at various sampling rates. Where results for a particular sampling rate are not provided in the table, the number of scan alerts was zero; these table rows have been omitted for brevity.

Table D.1 CAIDA Portsweeps at low sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	20	0	0.00%	20	0	0.00%	100.00%
1:5	1	0	0.00%	1	20	95.00%	5.00%

Table D.2 CRAWDAD Academic Hall Portsweeps at low sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	8221	0	0.00%	8221	0	0.00%	100.00%
1:5	1207	234	19.39%	973	7248	88.16%	11.84%
1:10	235	39	16.60%	196	8025	97.62%	2.38%
1:25	6	0	N/A	6	8215	99.93%	0.07%

Table D.3 CRAWDAD Residence Hall 100 Portsweeps at low sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	11882	0	0.00%	11882	0	0.00%	100.00%
1:5	194	28	14.43%	166	11854	98.60%	1.40%
1:10	21	1	4.76%	20	11881	99.83%	0.17%

Table D.4 CRAWDAD Residence Hall 13 Portsweeps at low sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	6507	6507	0.00%	6507	0	0.00%	100.00%
1:5	578	25	4.33%	553	6482	91.50%	8.50%
1:10	179	7	3.91%	172	6500	97.36%	2.64%
1:25	23	0	0.00%	23	6507	99.65%	0.35%

Table D.5 CAIDA Portscans at low sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	896	0	0.00%	896	0	0.00%	100.00%
1:5	190	147	77.37%	43	853	95.20%	4.80%
1:10	44	22	50.00%	22	874	97.54%	2.46%
1:25	19	4	21.05%	15	881	98.33%	1.67%
1:50	11	1	9.09%	10	886	98.88%	1.12%
1:75	1	0	N/A	1	895	99.89%	0.11%
1:100	1	0	N/A	1	895	99.89%	0.11%

Table D.6 CRAWDAD Academic Hall Portscans at low sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	105	0	0.00%	105	0	0.00%	100.00%
1:5	38	38	100.00%	0	67	100.00%	0.00%
1:10	4	3	75.00%	1	102	99.05%	0.95%

Table D.7 CRAWDAD Residence Hall 100 Portscans at low sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	403	0	0.00%	403	0	0.00%	100.00%
1:5	23	15	65.22%	8	388	98.01%	1.99%
1:10	1	1	100.00%	0	402	100.00%	0.00%

Table D.8 CRAWDAD Residence Hall 13 Portscans at low sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	263	0	0.00%	263	0	0.00%	100.00%
1:5	8	4	50.00%	4	259	98.48%	1.52%

Table D.9 CAIDA Portsweeps at medium sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	57641	0	0.00%	57641	0	0.00%	100.00%
1:5	7697	254	3.30%	7443	50198	87.09%	12.91%
1:10	2246	101	4.50%	2145	55496	96.28%	3.72%
1:25	559	20	3.58%	539	57102	99.06%	0.94%
1:50	261	12	4.60%	249	57392	99.57%	0.43%
1:75	168	4	2.38%	164	57477	99.72%	0.28%
1:100	112	3	2.68%	109	57532	99.81%	0.19%
1:200	34	1	2.94%	33	57608	99.94%	0.06%

Table D.10 CRAWDAD Academic Hall Portsweeps at medium sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	35911	0	0.00%	35911	0	0.00%	100.00%
1:5	32621	2965	8.26%	29656	6255	17.42%	82.58%
1:10	27533	367	1.33%	27166	8745	24.35%	75.65%
1:25	25624	38	0.15%	25586	10325	28.75%	71.25%
1:50	19117	11	0.06%	19106	16805	46.80%	53.20%
1:75	10750	4	0.04%	10746	25165	70.08%	29.92%
1:100	3363	6	0.18%	3357	32554	90.65%	9.35%
1:200	4	1	25.00%	3	35908	99.99%	0.01%
1:500	1	1	100.00%	0	35911	100.00%	0.00%

Table D.11 CRAWDAD Residence Hall 100 Portsweeps at medium sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	13930	0	0.00%	13930	0	0.00%	100.00%
1:5	1007	396	39.32%	611	13534	95.61%	4.39%
1:10	188	42	22.34%	146	13888	98.95%	1.05%
1:25	4	1	25.00%	3	13929	99.98%	0.02%
1:50	1	0	0.00%	1	13930	99.99%	0.01%

Table D.12 CRAWDAD Residence Hall 13 Portsweeps at medium sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	17958	0	0.00%	17958	0	0.00%	100.00%
1:5	1381	392	28.39%	989	35519	94.49%	5.51%
1:10	593	67	11.30%	526	35844	97.07%	2.93%
1:25	356	1	0.28%	355	35910	98.02%	1.98%
1:50	78	0	0.00%	78	35911	99.57%	0.43%
1:75	47	0	0.00%	47	35911	99.74%	0.26%
1:100	31	0	0.00%	31	35911	99.83%	0.17%
1:200	2	0	0.00%	2	35911	99.99%	0.01%

Table D.13 CAIDA Portscans at medium sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	5286	0	0.00%	5286	0	0.00%	100.00%
1:5	980	6	0.61%	974	4312	81.57%	18.43%
1:10	455	2	0.44%	453	4833	91.43%	8.57%
1:25	151	1	0.66%	150	5136	97.16%	2.84%
1:50	1	0	0.00%	1	5285	99.98%	0.02%

Table D.14 CRAWDAD Academic Hall Portscans at medium sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	6	0	0.00%	6	0	0.00%	100.00%
1:5	62	61	98.39%	1	5	83.33%	16.67%
1:10	1		0.00%	1	5	83.33%	16.67%

Table D.15 CRAWDAD Residence Hall 100 Portscans at medium sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	72	0	0.00%	72	0	0.00%	100.00%
1:5	9	8	88.89%	1	71	98.61%	1.39%
1:10	1	0	0.00%	1	71	98.61%	1.39%

Table D.16 CRAWDAD Residence Hall 13 Portscans at medium sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	39	0	0.00%	39	0	0.00%	100.00%
1:5	24	24	100.00%	0	39	100.00%	0.00%
1:10	2	2	100.00%	0	39	100.00%	0.00%

Table D.17 CAIDA Portsweeps at high sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	57898	0	0.00%	57898	0	0.00%	100.00%
1:5	10561	311	0.00%	10250	47648	82.30%	17.70%
1:10	4040	112	2.77%	3928	53970	93.22%	6.78%
1:25	1011	47	4.65%	964	56934	98.34%	1.66%
1:50	353	11	3.12%	342	57556	99.41%	0.59%
1:75	206	6	2.91%	200	57698	99.65%	0.35%
1:100	154	6	3.90%	148	57750	99.74%	0.26%
1:200	61	0	0.00%	61	57837	99.89%	0.11%
1:500	25	0	0.00%	25	57873	99.96%	0.04%
1:1000	2	0	0.00%	2	57896	100.00%	0.00%

Table D.18 CRAWDAD Academic Hall Portsweeps at high sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	31480	0	0.00%	31480	0	0.00%	100.00%
1:5	19124	704	3.68%	18420	13060	41.49%	58.51%
1:10	10458	252	2.41%	10206	21274	67.58%	32.42%
1:25	6194	32	0.52%	6162	25318	80.43%	19.57%
1:50	5333	10	0.19%	5323	26157	83.09%	16.91%
1:75	5060	9	0.18%	5051	26429	83.95%	16.05%
1:100	4799	5	0.10%	4794	26686	84.77%	15.23%
1:200	3787	3	0.08%	3784	27696	87.98%	12.02%
1:500	1202	0	0.00%	1202	30278	96.18%	3.82%
1:1000	2	0	0.00%	2	31478	99.99%	0.01%

Table D.19 CRAWDAD Residence Hall 100 Portsweeps at high sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	26565	0	0.00%	26565	0	0.00%	100.00%
1:5	2056	321	15.61%	1735	24830	92.34%	7.66%
1:10	612	106	17.32%	506	26059	97.77%	2.23%
1:25	82	18	21.95%	64	26501	99.72%	0.28%
1:50	10	2	20.00%	8	26557	99.96%	0.04%
1:75	6	2	33.33%	4	26561	99.98%	0.02%
1:100	3	2	66.67%	1	26564	100.00%	0.00%

Table D.20 CRAWDAD Residence Hall 13 Portsweeps at high sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	22649	0	0.00%	22649	0	0.00%	100.00%
1:5	1998	299	14.96%	1699	20950	92.50%	7.50%
1:10	750	90	12.00%	660	21989	97.09%	2.91%
1:25	189	17	8.99%	172	22477	99.24%	0.76%
1:50	93	4	4.30%	89	22560	99.61%	0.39%
1:75	73	2	2.74%	71	22578	99.69%	0.31%
1:100	63	2	3.17%	61	22588	99.73%	0.27%
1:200	42	1	2.38%	41	22608	99.82%	0.18%
1:500	4	0	0.00%	4	22645	99.98%	0.02%

Table D.21 CAIDA Portscans at high sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	6234	0	0.00%	6234	0	0.00%	100.00%
1:5	1277	43	3.37%	1234	5000	80.21%	19.79%
1:10	568	7	1.23%	561	5673	91.00%	9.00%
1:25	223	0	0.00%	223	6011	96.42%	3.58%
1:50	82	0	0.00%	82	6152	98.68%	1.32%
1:75	54	0	0.00%	54	6180	99.13%	0.87%
1:100	38	0	0.00%	38	6196	99.39%	0.61%

Table D.22 CRAWDAD Academic Hall Portscans at high sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	115	0	0.00%	115	0	0.00%	100.00%
1:5	182	172	94.51%	10	105	91.30%	8.70%
1:10	66	62	93.94%	4	6230	96.52%	3.48%
1:25	2	1	50.00%	1	6233	99.13%	0.87%

Table D.23 CRAWDAD Residence Hall 100 Portscans at high sense level.

Sampling rate	**# of scan alerts**	**# of *false scans***	***False scan rate***	***Detected scans***	**# of *missed scans***	***Missed scan rate***	***Consistency rate***
1:1	539	0	0.00%	539	0	0.00%	100.00%
1:5	77	53	68.83%	24	515	93.48%	6.52%
1:10	17	11	64.71%	6	533	98.37%	1.63%
1:25	2	2	100.00%	0	539	100.00%	0.00%

—This page intentionally left blank —

APPENDIX E: CODE FOR DETERMINING CONSISTENCY

```
#!/usr/bin/perl
#
# Elliot Proebstel
# Fall 2007
#
# This code looks for false positives in sampled alert data.
# It compares alerts raised on sampled data to alerts
# raised on the full trace.
# For an alert from the sampled data to "match" an alert
# from the full trace:
#
# Using sense_level "low", the alerts must match on:
# * Source IP address
# * Alert type
# * Time window (60 seconds)
#
# Using sense_level "medium", the alerts must match on:
# * Relevant IP address (sourceIP for all but "decoy" or
#   "distributed" scans)
# * Alert fields:
#   - (ICMP|TCP|UDP)
#    &&
#   - (Portscan|Portsweep|Sweep)
# * Time window (90 seconds)
#
# Using sense_level "high", the alerts must match on:
# * Relevant IP address (sourceIP for all but "decoy" or
#   "distributed" scans)
# * Alert fields:
#   - (ICMP|TCP|UDP)
#    &&
#   - (Portscan|Portsweep|Sweep)
# * Time window (600 seconds)
#
# The code tracks matches (a maximum of one match per sample-data alert),
# which it reports in a file that is cleverly named "matches", and false
# positives (alerts raised on sampled data that have no match in the full
# trace), which it reports in a file named "false_pos".
#
# The code requires a file called "all_portscans" to exist in the
# directory from which the script is run as well as in a
# directory accessible as ../full-trace
#

open(INPUT1,"<all_portscans"); # open "all_portscans" which has all
                               # portscan alerts
open(INPUT2,"<../full-trace/all_portscans"); # open "all_portscans"
                                             # in full trace
open(OUTPUT1,">matches");
open(OUTPUT2,">false_pos");

@alerts=<INPUT1>;
@baselines=<INPUT2>;
```

```
$totalMatches=0; # tracks total number of matches
$totalFPs=0;     # tracks total number of false positives

foreach $alert(@alerts) {

    $match=0;        # tracks current alert - matched yet or not

    @ids=split(/\}/,$alert);
    ($sourceIP, $destIP) = $ids[1] =~
m/(\d+\.\d+.\d+.\d+)\s+\-\>\s+(\d+\.\d+.\d+.\d+)/;
    $ids[0] =~ m/(\d\])(.*)(\[)/;
    $alertName=$2;
    if ($alertName =~ /(Distributed|Decoy)/) {$matchIP = $destIP;}
    else {$matchIP = $sourceIP;}
    $alertName =~ s/\(\S+\)//; # remove things in ()
    $alertName =~ s/\s+//;  # remove whitespace at the start
    ($alertProto, $alertType) = $alertName =~
m/(TCP|UDP|ICMP).+(Portsweep|Portscan|Sweep)/;
    ($aHours, $aMins, $aSecs) = $ids[0] =~ m/(\d+):(\d+):(\d+)/;
    $aTotalSecs=(($aHours*60*60)+($aMins*60)+$aSecs);

    foreach $base(@baselines) {

        if ($match==0)
      {
          ($bHours, $bMins, $bSecs) = $base =~ m/(\d+):(\d+):(\d+)/;
          $bTotalSecs=(($bHours*60*60)+($bMins*60)+$bSecs);
          $timeDiff=abs($bTotalSecs-$aTotalSecs);

          if (($base =~ /$matchIP/) &&
($base =~ /($alertProto).+($alertType)/) && ($timeDiff <= 600))
          {
            print OUTPUT1 "$alert $base\n";
            $match=1;
            $totalMatches++;
          }
      }
    }
    if ($match==0)
    {
      print OUTPUT2 "no match for $alert";
      $totalFPs++;
    }
}
print OUTPUT1 "Total number of matches: $totalMatches\n";
print OUTPUT2 "Total number of false positives: $totalFPs\n";

close(INPUT1);
close(INPUT2);
close(OUTPUT1);
close(OUTPUT2);
```