Urbashi Mitra, Antonio Ortega,
John Heidemann, and Christos Papadopoulos

Detecting and Identifying Malware: A New Signal Processing Goal

The Internet is now an essential part of many people's daily lives, both at home and at work. Today, it reaches all countries of the world, nearly a billion people, and supports trillions of dollars of transactions and millions of jobs. With this reach, it is not surprising that the Internet has been host to both positive and negative activities. Negative activities have grown from simple undesired use by unorganized hackers to extortion by organized crime [1]. The term *malware* has come to mean a range of harmful software that ends up on users' computers, such as keystroke loggers, Web monitors, viruses, worms, and zombie software allowing unauthorized remote control.

THE TERM *MALWARE* HAS COME TO MEAN A RANGE OF HARMFUL SOFTWARE THAT ENDS UP ON USERS' COMPUTERS, SUCH AS KEYSTROKE LOGGERS, WEB MONITORS, VIRUSES, WORMS, AND ZOMBIE SOFTWARE ALLOWING UNAUTHORIZED REMOTE CONTROL.

In what follows, we use the term *maltraffic* to include denial-of-service attacks, spyware reporting home, unauthorized applications (applications in violation of a company or Internet service provider's acceptable use policy, like peer-to-peer sharing, chat, and games), spam (both arriving and sending), worms, and similar kinds of network traffic. Some of these problems arise on compromised or misused internal computers, while others appear on zombies or compromised machines on the Internet to pummel trusted hosts. All are of growing concern.

While several technologies have been developed to counter threats and maltraffic, recent trends compromise many of today's defenses.

- An increase in encrypted traffic renders countermeasures based on content inspection ineffective.
- An increase of aggregation at network edges (due to caches, network-address-translation boxes, and other proxies) precludes filtering solely on IP addresses, as that could remove nonmaltraffic.
- An increase in traffic volumes means that maltraffic can easily hide in background traffic, yet still do great harm.
- The appearance of application cloaking, both intentional (varying port allocation or actively concealing application features) and unintentional (when applications are layered on existing protocols or systems use dynamic port allocation), means that simple approaches to identifying and filtering applications fail.

We suggest that these challenges present a unique opportunity to signal processing researchers. The notion of applying signal processing techniques to network analysis is not new, with prior success in the area of network tomography. However, we argue that the new challenges mentioned previously yield applications of signal processing that differ from those considered in network tomography [2]–[4]. Network tomography typically uses a limited number of active or passive measurements (typically at network edges) to infer network performance parameters and topology. The input signal typically consists of packet delays, round-trip times, loss, or similar features; tomography infers network characteristics using correlation techniques such as maximum likelihood estimation and Bayesian inference. In network tomography, multiple observation points may be required, and flows need to be separated from the aggregate. We propose to relax each of these requirements. Our goal is to collect data passively without separating traffic into flows and to design signal representations unique to the applications. To achieve this goal, we will use signal processing methods to characterize applications, not just network phenomena.

As in traditional signal processing problems, we take observations (in this case aggregate traffic traces) and from these extract relevant features and then apply detection techniques to determine whether maltraffic is present and, if possible, to identify those packets generated by malware. We believe that new methods in signal representation, classification and detection, transform domain techniques, and source separation are required for these applications, and these can potentially have a significant impact on countering the effects of malware on the Internet.

An understanding of the underlying network traffic is needed to determine which traffic features can be used to identify and detect maltraffic. Careful feature selection can minimize the effects of cloaking and encryption, since carefully selected features are impossible to cloak without reducing the maltraffic effectiveness. As an example, we have shown that the presence of denial-of-service attack packets can be detected using frequency domain features; concealing these features would require

1053-5888/06/$20.00©2006IEEE

reducing the attack rate, thus reducing attack effectiveness [5].

However, feature selection alone is not sufficient. An understanding of the measurement system and proper application of detection theory is needed to maximize detection sensitivity. Such approaches are essential to counter the challenges of edge aggregation and high traffic volumes, since potential maltraffic becomes a needle-in-the-haystack of large, aggregate, background traffic.

FEATURE EXTRACTION FOR MALTRAFFIC DETECTION

Applying signal processing to maltraffic detection involves 1) mapping observed network behavior to a signal representation and then 2) extracting features from this signal that can be used to to determine the presence of maltraffic. While detection techniques such as these have been used in the past in a networking context, we believe new methods have to be developed to address maltraffic. For example, our objective is to analyze aggregate traffic to make inferences about a particular outlier phenomenon. This makes it impractical to use techniques based on the analysis of adjacent packet interarrival times [6]. These become computationally infeasible for the networks of interest, as there will be many thousands of concurrent flows. Per-flow processing is very expensive, even on moderately large routers, and approaches such as statistical sampling [7] are incompatible with our goals of detecting small traffic exchanges and events in low-bandwidth flows.

Transform-domain techniques have been proposed recently as effective tools to extract underlying information from aggregate traffic. These approaches map relevant network traces into time series (e.g., number of packet arrivals per interval, packet arrival times) and then apply a transform to this time series signal. Intuitively, each transform domain sample can capture timing information corresponding to multiple packets. For example, if there is an underlying low-rate periodic stream in a trace, this will lead to energy appearing at the corresponding frequency in the transform domain. Examples of successful transform-domain techniques include use of the power spectral density [5], wavelets [8], and Lomb periodograms [9]. In each case, a standard analysis tool was used, with the hope that events to be detected would happen to produce a signature in the chosen representation domain.

Essentially, these techniques take off-the-shelf signal representation tools and combine them with equally well-known detection/classification tools. We believe that further improvements in detection performance can be achieved by taking a more formal approach to signal representation, feature selection, and detection. For example, one need not be bound by existing representation techniques; our goal should be to design new analysis and detection schemes where the transform domain technique development is performed jointly with the feature identification and modeling. This experience is borne out in other areas like speech recognition where, while many transforms have been considered [fast Fourier transform (FFT), cepstrum, and wavelets], it is the postprocessing coupled with the particular transform that determines the efficacy of the approach.

Our preliminary research points to several areas in which optimizing representation, transform, and detection for a given task have already shown some promise. As a first example, data obtained from real network measurements is often taken as if it represented an ideal measurement of underlying network behavior without considering that the measurement system itself may be introducing errors. Negligible or not, the measurement system must be modeled and analyzed to ensure reliability of the analysis. Furthermore, a careful understanding of measurement system limits allows operation near those limits, perhaps obtaining satisfactory results from an inexpensive measurement system. In the next section, we present an example of this need for a more formal approach to characterizing network measurement.

Second, applying off-the-shelf analysis tools (e.g., an FFT or various flavors of wavelet transform) directly to signals obtained from traces may be ineffective for two reasons. Signals derived from network data can be sparse in time (e.g., we have a nonzero signal only at times when packets arrive) and coarse in amplitude (e.g., integer valued, when a signal represents a number of packet arrivals). Both of these may pose problems for traditional approaches. We often know what type of signals are created by the phenomena that we try to detect (e.g., packets clocked at very regular intervals in the case of a bottleneck link). Thus, application-specific transforms that take into account the characteristics of signals and phenomena to be detected should be considered. As an example, this would argue for selecting transforms that map integers to integers and offer better resolution around those frequencies where these phenomena of interest are expected to manifest themselves.

Third, we believe that model-based detection offers promise. The difficulty of modeling general Internet traffic is widely recognized. However, our goal is fortunately much simpler: we need only model traffic characteristics that are relevant to the specific detection problem being considered. This approach is quite distinct from attempts to model general network traffic. In fact, the use of so-called inaccurate models has had a long and fruitful history in signal processing. An incomplete model can be effective if the mismatch is not related to the feature of interest.

APPROACHES TO APPLY SIGNAL PROCESSING TO MALTRAFFIC ANALYSIS AND DETECTION

We next consider two examples that highlight some of the challenges that arise when applying signal processing methods to this new area. We first examine the interplay between time series generation and the measurement system. We then evaluate how to define, catalog, and detect fingerprints that identify denial-of-service attacks.

UNDERSTANDING AND MODELING THE MEASUREMENT SYSTEM

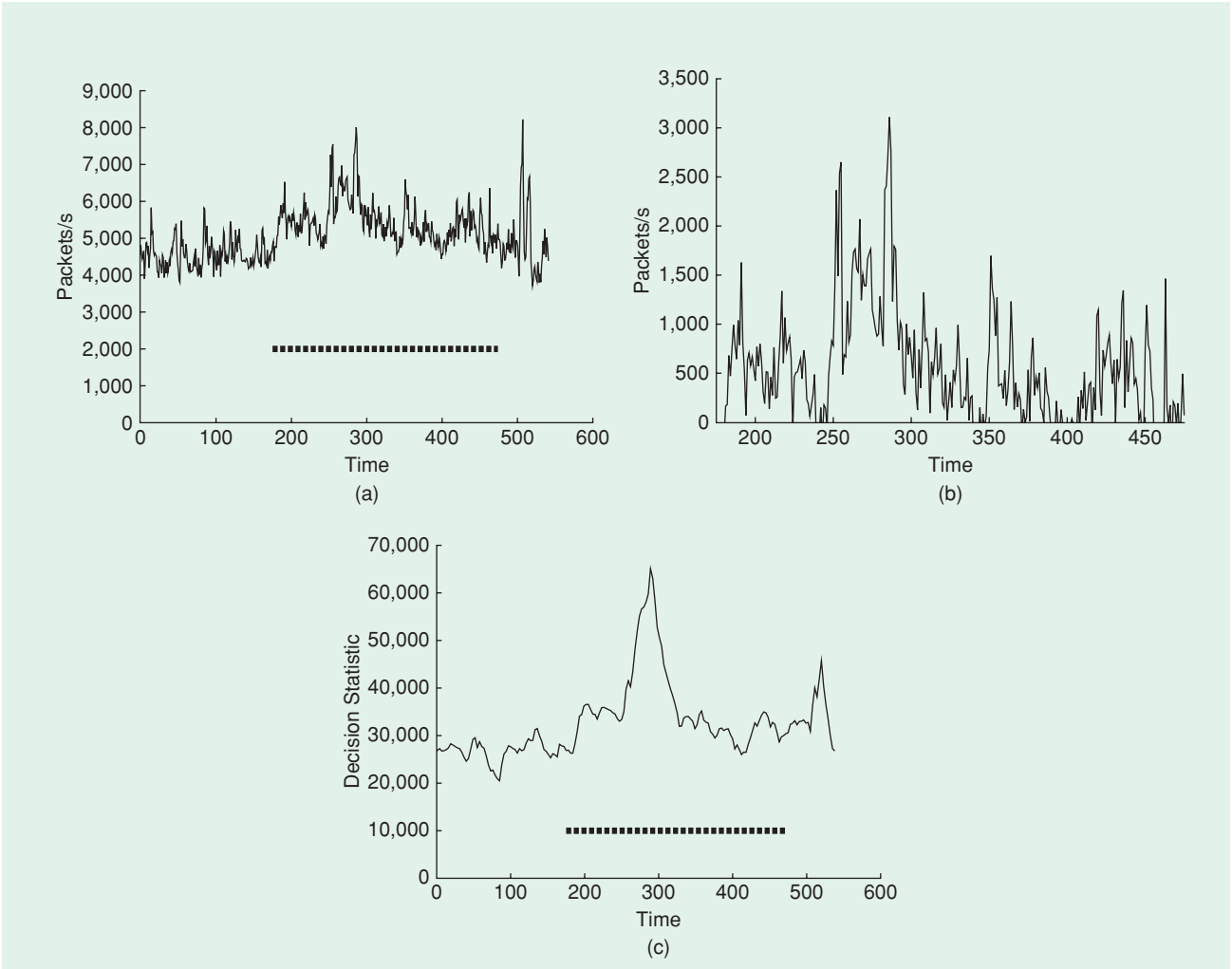
Our early trace measurement system used a stock PC, network card, and tcpdump to record packets. This system was

connected to the network either passively with an optical splitter or actively with port mirroring at a router. Let the link capacity be R b/s and represent packet arrivals as an ideal discrete (binary) signal, with sampling interval $1/R$ s. The board captures packets during a time interval, which we call the coalescing interval; all packets that arrive during one such interval are transferred to the host computer, which records their arrival using `tcpdump`. Thus, the timing produced by `tcpdump` is generated after information is received (in batches) and is affected in part by the processing speed of the host computer. The signal generated by `tcpdump` can be misleading; higher frequencies in this signal are likely to be generated by random variations in the `tcpdump` timing rather than by actual periodicities in packet arrivals in the network.

A better measurement approach would be to ignore the arrival times provided by `tcpdump` and instead analyze a signal where each sample corresponds to the total number of arrivals in each coalescing interval. This can be easily modeled using standard signal processing tools; the ideal binary signal is low-pass filtered and downsampled (corresponding to obtaining the total number of arrivals in each coalescing interval). Letting the discrete time Fourier transform (DTFT) of the original ideal signal be $X(e^{jw})$ and the coalescing interval be of length M samples, then the system we described is equivalent to first applying a filter $H(e^{jw}) = (1 - e^{-jwM})/(1 - e^{-jw})$ to $X(e^{jw})$ and then downsampling the results by a factor of M, so that the output of the measurement system has a DTFT

$$Y(e^{jw}) = \frac{1}{M}\sum_{k=0}^{M-1} X\left(e^{j(w-2\pi k)/M}\right) \times H\left(e^{j(w-2\pi k)/M}\right).$$

Note that with this explicit representation of the measurement system, we can both determine if the measurement system prevents us from observing a phenomenon of interest and predict the exact, observed, spectral signature, instead of its ideal signal on-the-wire, which will help in the detection process.

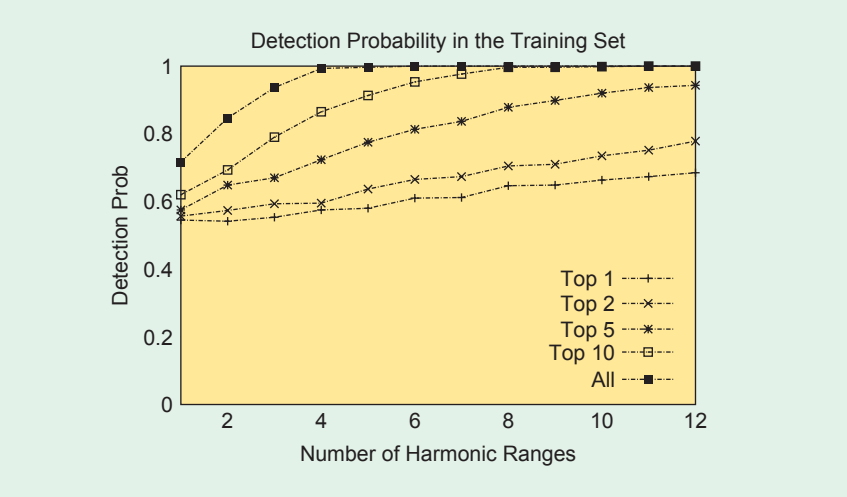


[FIG1] A low-volume attack of 670 packets per second is detected using change-point detection on the spectral content of the aggregate network traffic. (a) Aggregate network traffic in packets per second. Attack duration indicated by the line. (b) Attack traffic in packets per second. (c) Attack is detected using nonparametric detection techniques. Attack duration indicated by the line.

IDENTIFYING AND DETECTING ATTACK FINGERPRINTS

For a large class of problems stemming from malware, some form of detection or classification is inherent to the desired solution. Our initial approaches have focused on Fourier signal representations, which work well for certain applications, such as distributed denial-of-service (DDoS) attack classification and bottleneck detection. The fingerprinting of stationary DDoS attacks is based on the observation that spectral characteristics of an attack are hard to forge and attack scenarios appear to have unique spectral characteristics [10]. As such, we have successfully used maximum likelihood detection with Gaussian vector models to identify and thus classify repeated attacks [5], [10].

An important challenge is the presence of strong transients in the attack fingerprint, since such features are not well captured by a Fourier analysis. To overcome this shortcoming, we applied an initial transform method, the Mexican Hat wavelet, to 38 real-world DDoS attacks. We achieved a 90% attack detection rate with a negligible false positive rate. Figure 1(a) shows aggregate traffic seen in the network, in packets per second. Figure 1(b) shows the attack traffic alone: a TCP flooding attack of 670 packets/s lasting 175–475 s. With aggregate traffic, visual inspection of spectral data does not clearly indicate the presence of the attack. However, statistical detection techniques can find low-rate attacks reliably. Figure 1(c) shows that our preliminary decision statistic, computed based on a nonparametric change-point detector, increases rapidly during an attack.



[FIG2] Probability of detection as a function of the number of frequencies and harmonics considered. The x-axis shows how many harmonics are considered (1–12), while the several lines consider how many frequencies are considered in each harmonic (1, 2, 5, 10, or all). We consider the strongest *k* frequencies based on their ranked performance for the training set.

Another application we have explored is bottleneck traffic detection. Although not a maltraffic problem, per se, this study provides a methodology and intuition for more adversarial traffic. For moderate to low cross traffic, we have observed that a highly utilized transit or peering link will yield a strong frequency component proportional to the link speed and inversely proportional to the packet size yielding detection rates of 80% or higher, even when observed several hops away from the bottleneck. When mixed with aggregate traffic at higher rates, detection becomes more challenging. For example, there is no visual indication of 10 Mb/s bottleneck traffic in 50 Mb/s background traffic (on a 100 Mb/s aggregate link). More sophisticated approaches (multidimensional tests on the top frequencies) do improve performance at the expense of complexity as seen in Figure 2. The results of Figure 2 are achieved using a training set that is matched to the evaluation set by using data from the same time of day. If there is a mismatch between the training and the evaluation, we achieve very poor performance—around 50% detection rates. Thus for this problem, an important goal is to understand when training data provides a sufficiently good match to current conditions. While we recognize the difficulty in modeling general network traffic, the ability to match and model aspects of traffic specific to maltraffic detection can result in much better detection probabilities.

CONCLUSIONS

We believe that signal processing methods are powerful tools for defense against maltraffic on the Internet. While off-the-shelf techniques such as FFTs and popular wavelet families have been employed, precisely designed analysis and detection methods offer much more promise. A key challenge for this area of research is a lack of well-accepted models and even ground truth. However, such challenges exist as well in such arenas as speech and image processing. We conclude with three key observations.

- With signal processing techniques, we can detect phenomena using only packet timing information, i.e., without requiring inspection of packet contents, thus circumventing encryption.
- Using carefully designed measurement systems, we can apply detection theory to uncover the presence of very small signals of interest in aggregate traffic.
- By understanding the inherent patterns in traffic (who talks to whom and when) we can defeat cloaking by identifying behaviors inherent in the traffic patterns.

ACKNOWLEDGMENTS

The authors wish to thank Genevieve Bartlett, Xinming He, Alefiya Hussain, Wen-Tien Kung, Usman Riaz, and Rishi Sinha for their contributions to the results discussed herein. The authors also gratefully acknowledge the support of Cisco for research in this area.

AUTHORS

Urbashi Mitra (ubli@usc.edu) received the B.S. and the M.S. degrees from the University of California, Berkeley, and the Ph.D. from Princeton University. Since 2001, she has been with the Department of Electrical Engineering at the University of Southern California, Los Angeles, where she is currently a professor.

Antonio Ortega (ortega@sipi.usc.edu) received the telecommunications engineering degree from the Universidad Politecnica de Madrid, Spain and the Ph.D. in electrical engineering from Columbia University, New York. He is with the Department of Electrical Engineering, University of Southern California, where he is currently a professor.

John Heidemann (johnh@isi.edu) received his B.S. from the University of Nebraska-Lincoln and his M.S. and Ph.D. from the University of California, Los Angeles. He is a senior project leader at USC/ISI and a research associate professor at the University of Southern California in the computer science department.

Christos Papadopoulos (christos@isi.edu) received his Ph.D. in computer science in 1999 from Washington University in St. Louis, Missouri. He is an associate professor of computer science at Colorado State University.

REFERENCES

[1] J. Menn, "Deleting online extortion," *LA Times*, p. 1, Oct. 25, 2004.

[2] A. Adams, T. Bu, T. Friedman, J. Horowitz, D. Towsley, R. Caceres, N. Duffield, F.L. Presti, S.B. Moon, and V. Paxson, "The use of end-to-end multicast measurements for characterizing internal network behavior," *IEEE Commun. Mag.*, vol. 38, pp. 152–159, May 2000.

[3] M. Coates, M. Rabbat, and R. Nowak, "Merging logical toplogies using end-to-end measurements," in *Proc. ACM Internet Measurement Conf.*, Miami Beach, FL, Oct. 2003, pp. 192–203.

[4] Y. Zhang, M. Roughan, C. Lund, and D. Donoho, "An information-theoretic approach to traffic matrix estimation," in *Proc. ACM SIGCOMM Conf.*, Karlsruhe, Germany, Oct. 2003, pp. 301–312.

[5] A. Hussain, J. Heidemann, and C. Papadopoulos, "A framework for classifying denial of service attacks," in *Proc. ACM SIGCOMM Conf.*, Karlsruhe, Germany, Aug. 2003, pp. 99–110.

[6] D. Katabi and C. Blake, "Inferring congestion sharing and path characteristics from packet interarrival times," MIT, Laboratory Comput. Sci., Tech. Rep. MIT-LCS-TR-828, 2001.

[7] C. Estan, S. Savage, and G. Varghese, "Automatically inferring patterns of resource consumption in network traffic," in *Proc. ACM SIGCOMM Conf.*, Karlsruhe, Germany, Aug. 2003, pp. 137–149.

[8] P. Huang, A. Feldmann, and W. Willinger, "A non-intrusive, wavelet-based approach to detecting network performance problems," in *Proc. ACM SIGCOMM Internet Measurement Workshop*, San Francisco Bay Area, U.S.A., Nov. 2001, pp. 213–227.

[9] C. Partridge, D. Cousins, A.W. Jackson, R. Krishnan, T. Saxena, and W.T. Strayer, "Using signal processing to analyze wireless data traffic," in *Proc. ACM Workshop Wireless Security*, Atlanta, GA, U.S.A., Sept. 2002, pp. 67–76.

[10] A. Hussain, J. Heidemann, and C. Papadopoulos, "Identification of repeated denial of service attacks," in *Proc. IEEE Infocom*, Barcelona, Spain, 2006. SP