CyberSeer: 3D Audio-Visual Immersion for Network Security and Management

Christos Papadopoulos
University of Southern California
941 W37th Place
Los Angeles, CA 90089
+1-213-740-4780
christos@imsc.usc.edu

Chris Kyriakakis
University of Southern California
3740 McClintock Ave
Los Angeles, CA 90089
+1-213-740-8600
ckyriak@imsc.usc.edu

Alexander Sawchuk
University of Southern California
3740 McClintock Ave
Los Angeles, CA 90089
+1-213-740-4622
sawchuk@imsc.usc.edu

Xinming He
University of Southern California
941 W37th Place
Los Angeles, CA 90089
+1-213-740-6578
xhe@usc.edu

ABSTRACT

Large complex networks have become an inseparable part of modern society. However, very little has been done to develop tools to manage and ensure the security of such networks. Network operators continue to slave over endless daily logs and alerts in a struggle to keep networks operational. Perhaps the most formidable enemy of network operations today is the volume of management data that must be perused. Expensive commercial products attempt to visualize data but with limited utility, as witnessed by the prevailing use of command-line interfaces and homegrown scripts. In addition to data collection tools, operators need to immediately observe and debug the effects of their actions; yet that information is buried deep in the data that pours daily from monitoring equipment. Thus, they need better ways to abstract network events and better, more informative ways to render them.

In this work we first propose a new approach for abstracting network information, namely spectral representation. Second, we introduce immersive spatial audio representations of network events. Third, we introduce 3D interactive auto-stereoscopic (AS) displays for visual representations. We integrate the three techniques as follows. We use spectral techniques to extract complex events buried inside voluminous network traces and logs. Then, we create a desktop interactive immersive auto-stereoscopic 3D environment that is seamlessly integrated with multi-channel spatially rendered audio to render such events in a far more human-friendly fashion.

Categories and Subject Descriptors

C.2.3 [**Network Operations**]: Network management, Network monitoring, public networks.

General Terms

Management, Measurement, Documentation, Performance, Design, Reliability, Security, Human Factors.

VizSEC-DMSEC'04, October 29, 2004, Washington, DC, USA.

Keywords

Network visualization, network security, monitoring.

1. INTRODUCTION

Network management and security have become important tasks in protecting the cyber infrastructure. Networks are becoming more complex and maintenance requires a substantial effort from network operators. In addition, the Internet is plagued with viruses, trojans, portscans, DDoS attacks, and many other types of malicious activities. While network operators are performing admirably against such challenges, this comes at a steep price. Many human hours are wasted on everyday incidents, which must be examined, categorized, and perhaps acted upon. Fatigue becomes a major issue as network operators scour through endless logs, hundreds of daily emails and equipment alerts. Commercial products such as Intrusion Detection Systems (IDS) are costly and in general do not alleviate the problem, frequently adding to the deluge of alerts. Indeed, from discussions with our local ISP and information gleaned from mailing lists such as NANOG (North American Network Operators Group), it is clear that many network operators rely on open-source tools such as *tcpdump* [1], *tcptrace* [2], *snort* [3], etc., custom scripts, manual examination of logs and strategically located triggers for alerts of possible problems and/or malicious activity in the network.

Several visualization tools promised to simplify network administration, but with limited success. Such tools typically create graphical representations of the network with clickable links and nodes and typically return SNMP information from the various components. The tools can be programmed to generate alerts when certain parameters violate a preset threshold, e.g., when packet rate exceeds normal operating levels. A problem with such tools is lack of abstraction, as they understand very little of the semantics of traffic. For example, they cannot distinguish between web requests from a machine-generated attack and legitimate requests due to popular content. What is needed is a higher-level abstraction, which cannot only represent traffic parsimoniously, but is also capable of differentiating between various types of traffic and capture traffic behavior at different levels.

We make the important observation that much of network traffic exhibits periodic behavior. Such behavior ranges from periodic transmission of packets on a link, to protocol and

application behavior. We can parsimoniously characterize such periodicities in the frequency domain and build models based on such periodic behavior using spectral analysis techniques. Spectral techniques and tools are mature and have long been used in statistical analysis of periodic phenomena. Spectral analysis applied to network traffic may reveal several periodicities. For example, a protocol such as TCP exhibits periodicities due to its windowing behavior. Protocols such as BGP exchange regular messages every 30secs. A highly utilized link transmits packets periodically, governed by its speed and packet. Finally, many applications are inherently periodic, such as web requests by users, or continuous media applications such as audio and video. Spectral analysis may detect problems that often manifest themselves as abnormalities/disruptions to these periodic processes.

There are several early indications that spectral analysis may be applied to network traffic. Recently, it was shown that such techniques could be used in the study various phenomena, such as the composition of DDoS attacks [4], protocol behavior [5] and determining link characteristics [6]. Clearly, the ability to determine if "there are multiple sources in this attack", or "the window of this TCP flow is too small", or "this flow is coming from a DSL line or a cable modem" are examples of useful information that can be derived from spectral techniques. We believe that while demonstrating the applicability at various layers, the early work has only scratched the surface. Thus, the first important contribution of this work is the investigation of spectral techniques as a means to parsimoniously represent network traffic.

The second important contribution of this work is the investigation of multi-dimensional aural and visual representations of network traffic. Such techniques are capable of representing large amounts of data in natural ways that are easy for human operators to understand. Visual representations have been attempted before with various degrees of success. Many have resulted in little more than "eye candy" as the representations proved to be overwhelming, providing little useful information. A link turning red, for example is an important event, but not particularly useful unless accompanied by more relevant information, which inevitably sends operators back to scouring logs. Armed with models derived from spectral analysis, our work attempts to go beyond typical representations and investigate more meaningful representations such as ailing TCP flows, missed BGP updates, signs of overloaded links or identify machines that transmit at constant, high speeds.

Aural representations, while a powerful means for representing information for human consumption, have not been thoroughly investigated as a means to represent network traffic. Sound is a crucial, and often underestimated source of information about the physical world around us and provides input for many subconscious decisions. Often, we take actions based on aural input alone. Humans separate that sound from other background sounds in a few milliseconds and identify its direction to within a few degrees. In addition, we can perceive spectral changes (timbre), pitch changes, and variations in temporal patterns and perform grouping, which allows us to separate streams and relegate them to the background even if they consist of many individual different sounds. Our proposed work will capitalize on our ability to track a particular sound within other groups and bring it to the foreground selectively and allow events to be represented in a manner that does not demand the operator's constant, full attention. A second critical component is the temporal resolution of human hearing. Humans are capable of perceiving subtle rhythm changes for sounds that are repeating periodically. Our proposed sound synthesis models will be based on this knowledge and will use the time difference as a parameter to encode information about the data. Finally, aural representations are very memorable, which may enable an operator to "memorize' the status of the network.

While both aural and visual representations offer a natural interface to human operators, we will combine these modalities for registered spatial rendering of audio and video in multiple dimensions. For example, 3D sound can represent different physical parts of the network in space, allowing an operator to pinpoint a problem quickly. Integrating sound with a 3D visual representation allows a much larger visual field, which reduces the clutter of 2D dramatically. Such an environment allows the operator to truly immerse in the virtual network.

In summary, our work investigates data-driven models for network traffic that encode timing elements (from a few microseconds to several seconds) into 3D sounds and images to represent patterns of interest in network data. We pursue new algorithms for spatially registered audio and auto-stereoscopic 3D video rendering to represent patterns in multiple auditory and visual streams.

2. Spectral Techniques

We have argued that there exist several periodic phenomena in the Internet and spectral analysis can be used to characterize them. In this section we describe several examples of potential representations. These include, identifying a congested link, observing TCP behavior with small windows in large bandwidth-delay links, and observing the scheduling behavior of the operating system through network traffic.

2.1 Methodology

In order to apply spectral analysis, our data must be converted to a time-series. We create the time-series by sampling our data at regular intervals, and ensuring that the sampling rate is sufficient to capture the events of interest. For example, when sampling packet traces we use a tool such as *tcpdump* to capture and time-stamp packets and then sample the resulting trace to generate the time-series. Capturing and accurately time-stamping packet traces is perhaps the most demanding of our time-series generation process because we must ensure sufficient accuracy (often down to a microsecond). Our current trace infrastructure consists of high-end PCs running FreeBSD and the *NetGear GA620* Gigabit Ethernet card, for which open source drivers exist that are capable of performing partial packet transfers to reduce load on the PCI bus. We typically capture the first 68 bytes of each packet.

Creating time-series for other processes such as logs, SNMP data and *NetFlow* data is far less demanding in terms of time resolution, but demands time synchronization. We believe that NTP synchronization of the data collection points will be sufficient.

There have already been several results published about applications of spectral techniques. We give examples next.

2.2 Example 1: Visualizing a Single vs. Multi-Source DDoS Attacks

In Hussain et al. [4] it was demonstrated that there is a significant shift in the spectrum towards lower frequencies when multiple attack streams from different sources aggregate in a DDoS attack. Figure 1, which is taken from [4], shows the difference in spectrum between a single and multi-source attacks. These plots were obtained from real attacks captured at our monitoring point at our regional ISP.

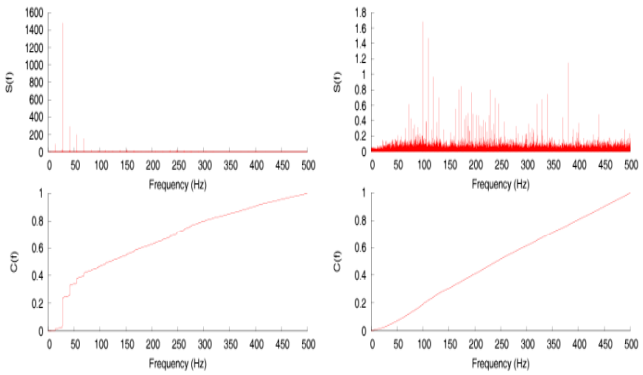

Figure 1: Multi-source vs. single-source DDoS attacks.

The figure shows a multi-source DDoS attack on the left. This attack was determined to contain multiple attackers through other means (in this case by detecting multiple sequence spaces in the IP ID field). The graph on the right was similarly determined to be a single-source attack. The spectrum plots at the top show a dramatic difference between the two attacks, which is also reflected in the cumulative normalized power spectrum plots at the bottom. This methodology can be applied towards the implementation of a tool that can quickly determine the number of sources in an attack, which will guide the operator in selecting means to neutralize the attack (for example, in selecting a traceback method).

2.3 Example 2: Visualizing a Congested Link

A highly utilized link (i.e., meaning queue length > 1) transmits packets at regular intervals depending on the link speed and packet size. Spectral techniques can be used to detect such frequencies in the packet stream. Next we describe a few experiments we carried out over our local network and over the wide-area Internet that demonstrate this technique.

In Figure 2 we see the spectrum of an un-congested and a congested link respectively. The experiment consists of a UDP flow generated by the *iperf* [7] tool between two PCs on a 100Mbps LAN segment. On the top figure we see the results of an experiment when the rate of the flow is artificially limited to well under 100 Mbps. This results in a spectrum where most if the frequency content is concentrated at the lower frequencies. On the lower figure we see a different experiment, where a flow is transmitting at near-full link rate. We see a clear dominant frequency at about 8 kHz, which is the rate 1500-byte packets are be transmitted on a 100 Mbps link.

The next experiment shows that the signature of a congested link can be detected in the spectrum of all background traffic. Here we have placed a trace machine on our ISP link to Internet II. The machine monitors both inbound and outbound traffic from all the customers of our ISP, including our university and a mix of other academic and commercial institutions. The average packet rate observed on this link is about 20,000 packets per second. The experiment involves a UDP flow from a machine outside our

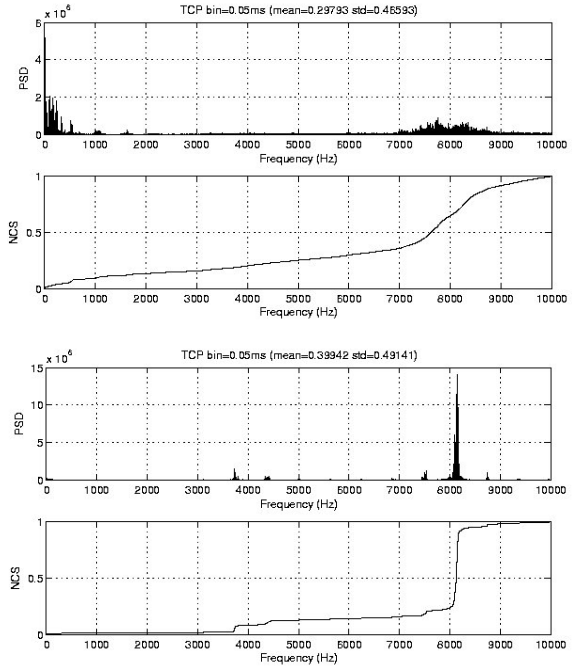

Figure 2: Spectrum of an un-congested link (top) and a congested link (bottom).

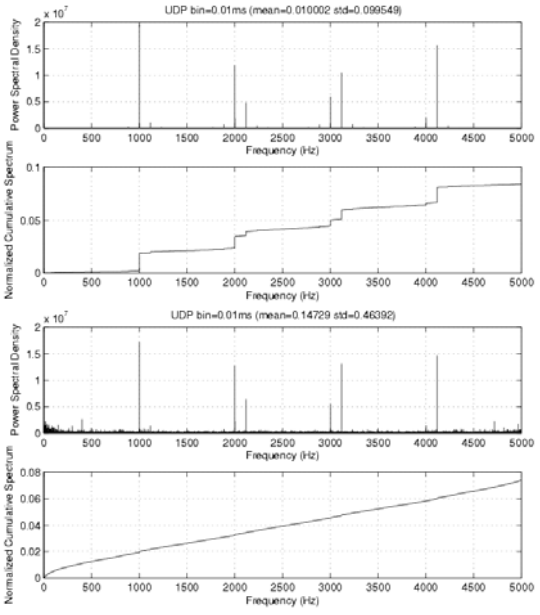

Figure 3: The spectrum of a UDP flow (top) is still visible in background traffic (bottom).

university that saturates the local access link and terminates to a machine inside our university. Our trace machine captures all the traffic on the Internet II link including our test flow. The packet rate of the test flow is about twenty times less that the average packet rate observed at this monitoring point.

Figure 3 shows the results. On the upper figure we observe the spectrum of our test flow, after filtering out of all other traffic. We can clearly see the signature of the overloaded link dominating the spectrum. On the bottom, we see the spectrum of all traffic captured. The signature of our test flow is clearly visible.

The above experiment was done over a real operational network and provides a strong indication that spectral techniques are capable of detecting signatures such as those produced by congested links. Such capability allows a network operator to quickly detect problems in the network and take corrective action. One example is a DDoS attack that saturates a link, which would generate a strong persistent spectrum.

2.4 Example 3: Observing Protocol Behavior

Several protocols exhibit periodic behavior. At the transport level, TCP employs a windowing mechanism for reliability and congestion control. Other protocols such as BGP, exchange periodic updates. HTTP transactions exhibit a periodic behavior based on user requests. Such behavior can be captured and characterized using spectral techniques. Spectral techniques can also help diagnose problems with such protocols. We describe a few examples next.

TCP employs a windowing mechanism, where acknowledgements from the receiver control the sender's transmission. The ability to monitor the behavior of the windowing mechanism in real-time, can go a long way towards determining the health of TCP flows. As an example we present the results from a cross-country experiment with TCP flows with socket buffers that are too small. The experiment involves running a TCP flow using *iperf* between two machines located on the east and west coast of the US. The default socket buffers allocated by *iperf* are too small to allow full utilization of the path, which results into throughputs of fewer than 10 Mbps in a 100 Mbps path Figure 4 shows the cumulative power spectrum and the power spectral density for this flow. There are two important points to observe. The first is that there is a clear signal present at 8 kHz, which corresponds to the frequency of 1500-byte packets the 100 Mbps path can support in back-to-back transmission. This is as expected. The second point, however, is a clear signal at a much lower frequency, which can be seem by zooming in at the bottom graph. This frequency coincides with the round-trip time of the flow, which is approximately 70 ms. This signal is produced by the windowing behavior of TCP. Since the socket buffers are too small to fill the pipe, TCP degrades to repeatedly sending maximum allowable size window of data with a period of one RTT, which is clearly captured in the figure. Therefore, our technique clearly detects a TCP flow that is not configured properly to utilize the available bandwidth.

In Figure 5 we show what happens when we gradually increase the socket buffers. As the window increases we notice that the signal at the link frequency becomes stronger, indicating that TCP better utilizes the available bandwidth. Finally, as the socket buffer approaches the bandwidth-delay product, the low frequency signal decreases dramatically, and the graph indicates a healthy TCP flow.

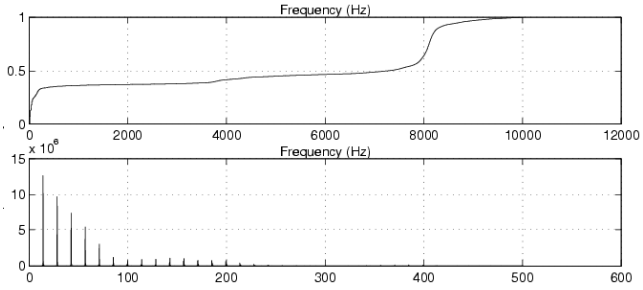

Figure 4: Observing TCP windowing and RTT behavior on a flow with a small window.

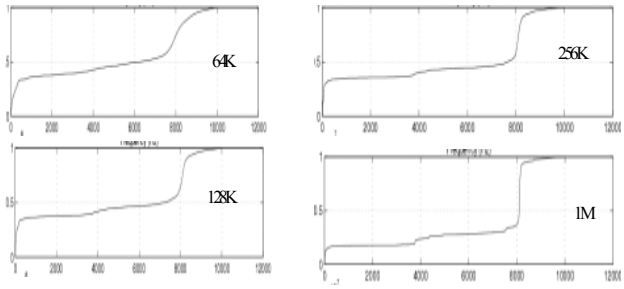

Figure 5: A TCP flow with various window sizes.

In addition to protocol behavior, spectral techniques can be used to characterize application layer behavior. For example, by monitoring traffic on a normal day (possibly for each web server) we can create a signature of a normal pattern of web requests. Then, if a server is attacked or "slash-doted", there will be a sudden increase in the pattern of web requests, which will be clearly visible in the spectrum.

Finally, we point out that the input data for spectral analysis need not come just from packet traces. Time-series data can be created from many other sources, such as SNMP traces, *NetFlow* logs, etc. Signatures could be constructed for several applications, such as RealAudio, VoIP, Mail, etc. The possibilities are virtually endless, but research is needed to determine which of these sources of data are more interesting and how to best represent them for the human operator.

3. Immersive Sound for Network Data Sonification

Sonification research has focused on the use of sound to present information about complex data. Previous work in this area can be divided into two broad categories: (i) data-driven models for sound synthesis and (ii) auditory displays for presentation of such data.

In this work we investigate elements in both areas as they apply to the network monitoring, management and attack problems. One of the novelties of our approach lies in the fact that we explore the spatial dimension for rendering sound as an additional variable that can be used to convey information about the data. A second novel element is the use of Gaussian Mixture Models (GMMs) to synthesize sounds in response to changes in data patterns. Finally, we combine these techniques with spatially registered

auto-stereoscopic (no-glasses) graphics, as described later, to create an immersive audio-visual interaction space for users. To our knowledge, none of these approaches has been employed in the field of data sonification.

Sound is a crucial, and often underestimated source of information about the physical world around us and we have learned to use it to make subconscious decisions. Many times these decisions involve actions that are taken with no visual input. For example, the sound of an approaching car behind us alerts us to move even if we are engaged in conversation or performing some other task. In a few milliseconds we can separate that sound from other background sounds and identify its direction to within a few degrees. This ability to localize is based on background processing that translates differences in both loudness and time of arrival of sound at the two ears into direction.

In addition to our ability to perform precise localization even in noisy environments, we can perceive spectral changes (timbre), pitch changes, and variations in temporal patterns. A higher order function performed by human listeners is that of grouping. This is particularly important in multi-stream sonic environments because it allows us to separate streams based on their relative grouping. Thus, groups of sounds can be relegated to the background even if they consist of many individual different sounds. Our ability to notice or track a particular sound within those groups and bring it to the foreground selectively is central to the proposed work. The multi-channel audio rendering system that we investigate is be capable of rendering such groups in a way that allows the operator to monitor them in the background while focusing on visual tasks.

A second critical component is the temporal resolution of human hearing. Sound that arrives at each ear with a time difference of a few microseconds can give information to the listener as to its direction of arrival. As the time difference increases to a few milliseconds the sound begins to be perceived as two individual sound sources. Increasing the difference in arrival time to above 25 milliseconds allows us to perceive subtle rhythm changes for sounds that are repeating periodically. This ability to notice fine differences in rhythm changes is maintained as the arrival time difference grows to about 1 second. Our sound synthesis models are based on this knowledge and will use the time difference as a parameter to encode information about the data.

In summary, we investigate data-driven models for network traffic that encode these timing elements (from a few microseconds to several seconds) into sounds that can be used to identify patterns of interest in network data. We are developing new algorithms for spatial audio rendering that will allow us to represent patterns in multiple auditory streams. Finally, we will perform extensive subjective evaluations of our spatial sonification methods to identify the auditory parameters that provide the greatest improvement in comprehension and detection of network traffic problems.

3.1 Advantages of Auditory Displays

Kramer [8] summarizes the benefits of auditory displays as (i) eyes-free use, (ii) rapid detection of acoustic signals, (iii) alerting, (iv) orienting, (v) backgrounding, (vi) parallel listening, (vii) acute temporal resolution, (viii) affective response, and (ix) auditory gestalt formation.

In the network traffic domain, eyes-free displays are critical because the operator is required to maintain visual contact with a very large number of elements on one or more displays. Research has shown [9] that the human response time to an audio signal can be significantly shorter than to a visual signal. This capacity for rapid detection indicates that audio signals can be crucial in the design of interfaces to monitor and identify sudden changes in network data.

Alerting using sound has been used in numerous applications. In this work we will combine the benefits of alert sounds with spatialization to provide rapid orienting so that operators can quickly focus their attention in the direction that is required on a wide-field-of-view display. This is a case of "the ears telling the eyes where to look" and there is evidence in the literature to suggest that visual search performance is enhanced when augmented by sound [10].

The backgrounding capability of the human auditory system can relegate some sounds to a lower priority while still maintaining sufficient awareness so that any change will draw the operator's attention. This allows the user to increase the amount data streams that are monitored and yet have an enhanced ability to identify relevant events. When combined with our ability to perform parallel listening to keep track of several concurrent sounds, the operator can analyze and explore multiple data sets simultaneously in order to search for correlations that may otherwise be overlooked.

The acute temporal resolution that human listeners exhibit is a critical component that provides the capability to monitor time-sequenced data that covers a very large range without loss of resolution.

An area that has been largely ignored is that of affective response to sounds. We will build on our extensive experience in sound design to create displays that are engaging and capable of conveying subtle information through sound. We know from our own experience that sound can suggest emotions such as fear, anticipation, calm, imminent threat, humor, etc. The challenge here will be to derive models that can create such suggestions based on particular events in the data. Finally, auditory gestalt formation allows us to understand overall relationships and trends in large data sets [11]. These gestalts allow the user to identify meaningful events in a data stream.

3.2 Representing Data with Sound

The fundamental challenge to be addressed here is to find a set of truly useful auditory parameters that can be linked in meaningful ways to network data so that a measurable enhancement in monitoring performance can be achieved.

Previous work in this area has been limited to variations in the parameters provided by the Musical Instrument Digital Interface (MIDI). These systems provide the capability to play musical notes in a variety of "instruments" by controlling pitch and loudness. They provide no control over other useful parameters such as rates of attack and decay and time varying spectral changes. Furthermore, MIDI systems are limited to 7 bit words thus severely limiting their usefulness in representing subtle changes in data linked to sound.

In our sound-synthesis aspect of this work we will focus on sound generation methods that can create rich sonic textures with a

much higher degree of resolution than what MIDI systems provide.

There are several approaches that will be considered as candidates for sound synthesis. The first will be that of parameter mapping in which incoming network flows will be mapped to a specific parameter of a sound synthesizer. These mappings can be classified as 0th-order in which the data stream itself is converted into a sequence of audio samples and listened to [12]. The data itself is shifted into the audible frequency domain (20 Hz to 20 kHz). Digital filtering techniques such as matched filters will be explored for performing frequency selective operations and identifying the presence of various periodic components in the data while suppressing unwanted noise (Figure 6). The 0th-order approach will be compared to 1st-order mapping in which the incoming data controls a parameter of the synthesis model (e.g., the amplitude of an oscillator, or in our case the means of the Gaussian Mixture Models described below).

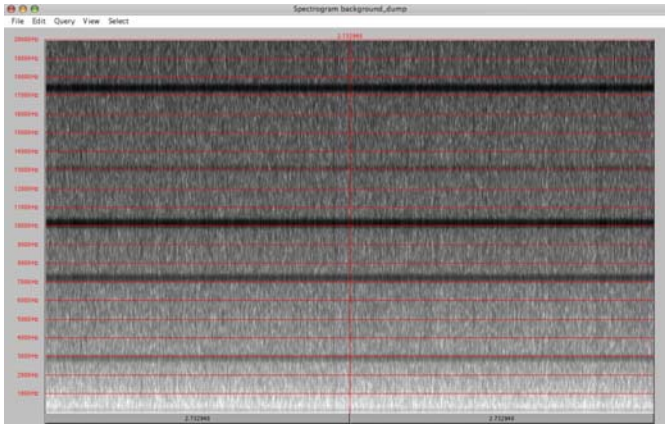

Figure 6: A time (horizontal axis) - frequency (vertical axis) representation of network traffic data shows strong periodicities in the signal (as indicated by the dark horizontal lines at specific frequencies). Matched filtering methods will allow us to improve the signal-to-noise ratio between these components of interest and the background noise.

One of the key problems associated with network traffic is the sampling rate. Meaningful sampling rates in this domain do not necessarily correspond to sampling rates optimal for high quality audio. A common approach to this problem is to simply re-sample or time compress (or time expand) the data. This, however, causes changes in the length of events in the data that may be missed due to being too fast or too slow to detect in an audio signal. Alternative methods of using a carrier signal to modulate the data into an FM signal cause additive shifts in frequencies that do not preserve the harmonic relationships. Most of the sound synthesis models that have been investigated rely on sinusoidal sound synthesis and are thus susceptible to the problems described above.

We propose a novel approach, based on techniques originating from the field of speech synthesis. Our methods will be based on a particular area of speech synthesis called voice conversion, in which the objective is to modify a speech waveform so that the context remains as is but appears to be spoken by a different speaker [13]. This is achieved by modifying the short-term spectral envelope using functions that are obtained during a training phase. In previous related work [14] we synthesized music signals by using the reference residual and the cepstral coefficients in the ideal case where the desired sequence of cepstral coefficients was correctly predicted. The result was a synthesized signal that had all of the desired characteristics.

In this work we seek to implement a model that synthesizes sounds in response to changes in data patterns. We will start with a set of sounds that are created deterministically from known data patterns and then use Gaussian Mixture Modeling (GMM) to represent the sounds. GMM methods model a signal as the realization of a probability density function that is a mixture of Gaussian distributions, with the parameters (means, covariance matrices, and prior probabilities) determined during this training phase. The challenge here will be to apply the GMM models to new data patterns and create sounds that appropriately represent those changes. Our research will focus on methods for estimating the appropriate GMM parameters based on what we know from the GMM models created during the training phase.

This parametric synthesis approach will not suffer from the time compression-expansion problems of sinusoidal methods and is expected to be less computationally intensive as well.

3.3 Metaphorical and Affective Sound Synthesis

One of the expected difficulties in the parameter mapping approach described above is that the resulting sounds are likely to be highly dissonant and thus difficult to listen to for extended periods of time. Furthermore, it may prove difficult to train operators to listen for a particular type of sound.

To address these problems we are, in parallel, investigating sound synthesis methods that use metaphorical and affective associations. Metaphorical associations are those that use a real world phenomenon to associate with changes in the data. For example, the sound of a beaker filling with water can be used to sonify the progress bar that moves during a file download. Affective associations are used to link a user's learned experiences with trends in the data. For example, most western hemisphere users will associate a sound similar to the theme from the movie "Jaws" with some type of imminent threat.

Since many of the decisions that a network operator needs to make when trying to determine whether a particular packet pattern represents an attack relate to "quantity" or "rate", it is useful to study synthesis methods for sound that convey these effects. Through subjective listening evaluations we are investigating whether the notion of "more" is best conveyed through increased *loudness*, *brightness* (high frequency boost), *speed*, and *pitch shifting*.

In addition to these metaphorical associations, we are investigating the effectiveness of affective associations. The idea here is to degrade some aspect of an otherwise "pleasant" or harmonic sound in response to undesirable changes in the data. For example, *dissonance* can be created by adding non-harmonic partials at a rate proportional to the rate of change in the data being monitored. Another type of variation in the data can cause *spectral changes* that include reduction or elimination of spectral components in a full-spectrum signal (*e.g.,* an orchestral classical music piece).

Several interesting research questions arise with these types of metaphorical and affective associations. We are examining cases

in which metaphorical and affective changes augment each other. This could happen, for example, when the metaphorical cue of more brightness indicates an undesired increase in packet rate and is then augmented by the affective cue of an undesirable spectral change. On the other hand, if the metaphorical cue of pitch shifting to a lower frequency indicates a desired data pattern, but the affective cue of lower frequencies indicates an imminent threat then there is a subjective conflict that must be resolved. These relationships between the two types of cues and the generalization of such cues across large user groups are a key component of this work and will be addressed with subjective criteria that are an extension of our previous work in the assessment of perceived quality in audio systems [15][16].

3.4 Multi-Channel Rendering of Network Traffic Sounds

In previous psychophysical research it has been shown that the human auditory system has the ability to parse, monitor, and alternate attention among multiple sound streams when these streams are rendered at different locations in 3D space [17]. This improvement in localization and comprehension has been studied for the case of speech signals [18].

We are investigating the advantages of spatially rendered audio for sonification. To achieve this, we build on our previous work of multi-channel audio rendering for immersive applications [19][20].

Traditional two-channel stereo is limited in its ability to render sound in 3D space because sonic images can only be rendered between the two loudspeakers and are accurately perceived only if the listener is seated at the exact center. Multi-channel audio systems have two key advantages: (i) they can render sound at any location in 3D space and (ii) the sensitivity to listener position decreases linearly as the number of loudspeakers increases [21].

Multi-channel (surround sound) systems have been developed that use three front channels and two surround channels. These systems were designed for watching movies on a screen in front of the audience and with no provision for precise sound localization outside of the screen boundaries. The two additional surround channels render ambient sound that is specifically designed to avoid localization so as to avoid diverting the attention from the action on the screen.

In our proposed multi-channel system, the design requirements are that sound must be precisely localized by multiple people simultaneously. This system will be integrated with stereoscopic display technology and, therefore, it is critical to ensure that the spatial relationships between visual objects and sound associated with them are accurately maintained. Because psychoacoustics tells us that humans are able to resolve direction of sound better in the front hemisphere [22], we increase the number of front channels from three to five. The two additional channels are placed at approximately ±55° with respect to the center. It has been shown [23] that this is the preferred direction of arrival for side reflections in order to enhance the perception of space. A rear center surround channel will be added to fill in the gap between the two surround loudspeakers. We introduce elevation information through two height channels placed above the front left and right loudspeakers. Early experiments that we have performed with this configuration have shown that these 10 channels significantly increase the sense of localization and envelopment for a variety of program material.

There are two key research challenges that we address. The first focuses on what we call "volumetric audio rendering". While multi-channel surround sound systems, such as the one we are proposing, can provide good envelopment and localization, they are not capable of providing distance cues. This is something that well-designed headphone systems utilizing head-related transfer functions can do well. We are investigating novel ways to render HRTF-based sound over multiple loudspeakers in such a way that multiple listeners in the same room can experience it simultaneously and also perceive distance cues. This involves investigation of new methods of crosstalk cancellation that allows HRTF-based audio intended for headphones to be rendered over multiple loudspeakers.

Our work involves experimentation with several types of adaptive algorithms to determine the weighting vectors for the filters required to perform the spatialization of sound as well as those required to cancel the undesired crosstalk terms that are present in all loudspeaker-based systems. The convergence speed of traditional adaptive filter design algorithms is too slow for real time operation. We are investigating algorithms such as the Karhunen-Loeve Transform (KLT) to de-correlate the inputs by preprocessing them with a transformation that is independent of the input signal [15]. Multi-rate critical band filters will be used to improve performance while at the same time minimizing computational complexity.

3.5 3D Visualization

The temporal and spatial bandwidth of vision is much greater than the other senses combined, and more than half of the brain is dedicated to visual information processing [25][26]. Humans can absorb and perceive large amounts of visual information, particularly when the 3D senses enabled by binocular vision are active. Visualization has been used for many years to take multidimensional data and format it into a form suitable for simplified human interpretation and analysis. Although there have been several studies on visualization techniques for network analysis, there has been little work on these techniques for intrusion detection. Varner [27]summarizes these efforts. None of them have used an auto-stereoscopic display combined with immersive spatial audio. Becker et al. [28]describe several graphical techniques for representing network data: link maps on a 2D topology that display parameters as lines, colors, textures, etc. Eick [29] describes a hierarchical structure for network visualization in 3D. Koutsofios et al. [30] describe a data exploration and visualization system for analyzing telecommunications networks. Cox and Eick [31] explore 3D visualization to help overcome human observer problems with information clutter. Estrin et al. describe a network visualization tool [32] having time plots of events, packet-level animation, protocol graphs, editing tools and scenario editing under control of a multi-window user interface. Grinstein et al [33] describe one of the few studies on graphical analysis of network intrusions and make one of the first comparisons on the effectiveness of visualization techniques for various kinds of data.

3.6 3D Auto-stereoscopic (AS) Displays

Very recently, new types of single and multiple viewer high-quality auto-stereoscopic (AS) display systems have become

available. These displays *do not* require the use of head-tracking, external glasses or goggles, and make the image and video viewing experience more natural and less fatiguing. This glasses-free feature has tremendous potential advantage in this project for improving the immersive experience. These displays use liquid crystal (LC) or plasma flat-panel technology, and are currently available in sizes exceeding 50".

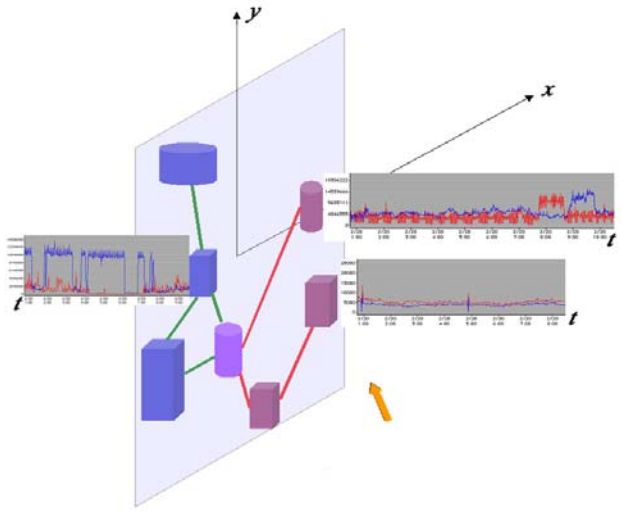

Figure 7: 3D oblique display with time history of packet flows in and out of selected nodes as red and blue time functions on the *z-axis*. Spatial audio is heard in register at locations of interest.

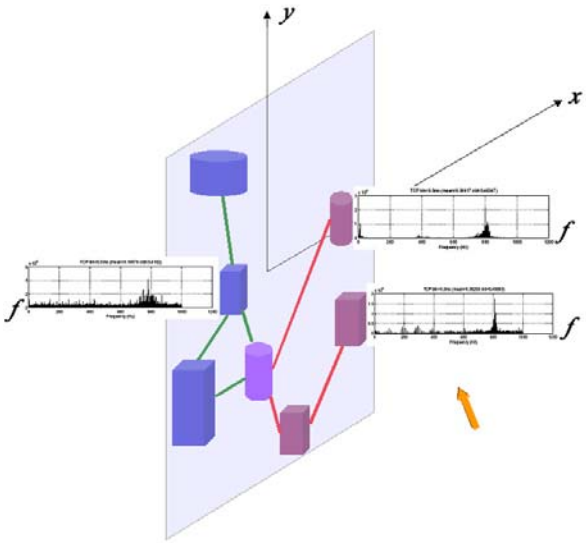

Figure 8: 3D oblique display with traffic power spectrum at selected nodes as red and blue time functions on the *z-axis*.

4. Immersive Examples of Operational Scenarios

Figure 7 and Figure 8 show various examples of 3D auto-stereoscopic visualization in which different information is superimposed on an (x, y) mapping of network topology. Figure 7 shows the time history of packet flows in and out of selected nodes as red and blue time functions on the *z-axis*. Automatic alerts activate audio signals and/or flow displays, or they are selected by user interaction. The user can vary the time scale of the alerts or displayed data, link it to other database information and manipulate the data in the spatial audio and visual domain. Figure 8 shows power spectra on the z-axis at various frequency scales for various nodes. The general audio and video rendering tools developed enable many different modes of detailed analysis. For example one or more power spectra as in Figure 8 can be mapped into spatial audio in which zero or low frequency information is located in front of and a few meters to the left of the user, while high frequency information is located in front of and a few meters to the right.

Our vision of an immersive auto-stereoscopic 3D display that is completely integrated with multi-channel immersive sound is shown in Figure 9. A set of loudspeakers provides the audio that is in spatial register with alerts and information on the display. The audio alerts map time or spectral information over the full spatial extent of the display or over broader regions extending to full 360-degree coverage as depicted. The user interacts with the displayed information by means of a 3D cursor system that uses a small light pen that is tracked by small video cameras (not shown in Figure 9) [34].

The parameters or features that are selected and mapped to the 3D audio-visual space are extremely general. Information from detection and analysis tools we describe can be combined with other data sources and mapped singly or combination. The overall objective is to augment the cognition process, enhance and expand the immersive analysis tools for users.

Figure 9: An immersive auto-stereoscopic 3D video and audio environment.

5. Conclusions

In this work we presented examples of how to abstract network information using spectral representation. Then, we proposed to use immersive spatial audio representations of network events and introduced 3D interactive auto-stereoscopic (AS) displays to handle the large amount of data generated by traffic traces, logs and events generated by intrusion detection and network management systems. Our goal is to create an interactive and immersive auto-stereoscopic 3D environment to render such voluminous data in a far more human-friendly fashion.

6. REFERENCES

[1] http://www.tcpdump.org/

[2] http://www.tcptrace.org/

[3] http://www.snort.org/

[4] A. Hussain, J. Heidemann and C. Papadopoulos, “A Framework for Classifying Denial of Service Attacks.” In Proceedings of Sigcomm 2003, Karlsruhe, Germany, August 2003.

[5] A. Broido, E. Nemeth and K.C. Claffy, “Spectroscopy of DNS Update Traffic,” Presented at Sigmetrics 2002.

[6] A. Broido, R. King, E. Nemeth and K.C. Claffy, “Radon Spectroscopy of Inter-Packet Delay,” High Speed Networking (HSN) 2003 Workshop, 2003.

[7] http://dast.nlanr.net/Projects/Iperf/

[8] G. Kramer, “An Introduction to Auditory Display”, In Auditory Display, edited by G. Kramer, SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII, Addison-Wesley, 1994.

[9] G.H. Mowbry and J. W. Gebbard, “Man’s Senses as Informational Channels.” In Human Factors in the Design and Use of Control Systems, edited by H. W. Sinaiko, 115-149, New York: Dover, 1961.

[10] D.R. Perrott, K. Saberi, K. Brown, and T.Z. Strybel, “Auditory Psychomotor Coordination and Visual Search Performance”, Perception and Psychology, 48, 214-226, (1990).

[11] S.M. Williams, “Perceptual Principles in Sound Grouping”, In Auditory Display, edited by G. Kramer, SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII, Addison-Wesley, 1994.

[12] G. Kramer and S. Ellison, “Audification: The Use of Sound to Display Multivariate Data”, In Proceedings of the International Computer Music Conference (ICMA), 214-221, San Francisco, CA 1991.

[13] G. Baudoin and Y. Stylianou, “On the transformation of the speech spectrum for voice conversion”, in IEEE Proc. Int. Conf. Spoken Language Processing (ICSLP), Philadelphia, PA, pp. 1405-1408, October 1996.

[14] A. Mouchtaris, S. S. Narayanan, and C. Kyriakakis, Multichannel Audio Synthesis by Subband-Based Spectral Conversion and Parameter Adaptation, accepted for publication IEEE Trans. Speech and Audio Processing.

[15] D. Yang, H. Ai, C. Kyriakakis and C.-C. Kuo, “High Fidelity Multichannel Audio Coding with Karhunen-Loeve Transform”, IEEE Trans. on Speech and Audio Processing, Vol. 11, No. 4, July 2003.

[16] D. Yang, H. Ai, C. Kyriakakis and C.-C. Kuo, “Progressive Syntax-Rich Coding of Multichannel Audio Sources”, EURASIP Journal on Applied Signal Processing, Vol. 2003, No. 10, Sept. 2003.

[17] A.S. Bregman, “Auditory Scene Analysis”, MIT Press, Cambridge, MA, 1983.

[18] A.W. Bronkhorst and R. Plomp, “The Effect of Head-Induced Interaural Time and Level Differences on Speech Intelligibility in Noise”, J. Acoust. Soc. Am. 83, 1508-1516, 1988.

[19] A. Mouchtaris, P. Reveliotis, and C. Kyriakakis, “Inverse Filter Design for Immersive Audio Rendering over Loudspeakers”, IEEE Transactions on Multimedia, 2(2), 77-87, (2000).

[20] A. Mouchtaris, S. S. Narayanan, and C. Kyriakakis, “Virtual Microphones for Multichannel Audio Resynthesis”, EURASIP Journal on Applied Signal Processing (JASP), Special Issue on Digital Audio for Multimedia Communications, 10, 968-979, 2003.

[21] G. Theile and G. Plenge, “Localization of Lateral Phantom Sources,” Journal of the Audio Engineering Society, vol. 25, pp. 196-199, 1977.

[22] J. Blauert, “Spatial Hearing: The Psychophysics of Human Sound Localization”. MIT Press, Cambridge, MA. 1983.

[23] Y. Ando, Concert Hall Acoustics, Springer-Verlag, New York, 1985.

[24] R. Madigan and D. Williams, “Maximum-Likelihood Procedures in Two Alternative Forced-Choice: Evaluation and Recommendations”, Perceptual Psychophysics, 42, 240-249, 1987.

[25] R.M. Friedhoff and M.S. Peercy, Visual Computing. Scientific American Library, New York, 2000.

[26] D. Marr. Vision, W.H. Freeman and Co., 1982.

[27] P. Varner, Security, Sonification, and Visualization: A State-of-the-Art Report, http://dependability.cs.virginia.edu/bibliography/state-of-the-art.pdf.

[28] R.A. Becker, S.G. Eick, and A.R. Wilks, "Visualizing Network Data," IEEE Transactions on Visualization and Computer Graphics, 1(1):16–28, 1995.

[29] T. He and S.G. Eick, "Constructing interactive network visual interfaces,"Bell Labs Technical Journal, 3(2):47–57, April-June 1998.

[30] E.E. Koutsofios, S.C. North, R. Truscott, and D. A. Keim, "Visualizing large-scale telecommunication networks and services," in Proceedings of the IEEE Visualization 97 Conference, pages 457–461, San Francisco, CA, 1999.

[31] K.C. Cox, S.G. Eick, and T. He, "3d geographic network displays, "SIGMOD Record, 25(4):50–54, 1996.

[32] D. Estrin, M. Handley, J. Heidermann, S. McCanne, Y. Xu, , and H. Yu., "Network visualization with NAM, the Vint Network Animator," IEEE Computer, 33(11):63–68, November 2000.

[33] G. Grinstein, "Workshop on information exploration shootout project and benchmark data sets: Evaluating how visualization does in analyzing real world data analysis problems, "In Proc. of the IEEE Visualization 97 Conference, pages 511–513, Phoenix, AZ, 1997.

[34] Z.Y. Alpaslan and A.A. Sawchuk, "Three-Dimensional Interaction with Autostereoscopic Displays," *Proc. Stereoscopic Displays and Virtual Reality Systems XI Symposium, Proc. SPIE*, Vol. 5291, San Jose, CA, 2004.