

# Evaluating the Impact of Network I/O on Ultra-Low Delay Packet Switching

George Baltas and George Xylomenos  
 Mobile Multimedia Laboratory, Department of Informatics  
 Athens University of Economics and Business  
 Athens 10434, Greece  
 gbaltas@aueb.gr, xgeorge@aueb.gr

**Abstract**—Low latency is a crucial requirement for demanding conferencing applications, such as Networked Music Performance (NMP), the collaboration of musicians in real time. Modern conferencing systems employ a Selective Forwarding Unit (SFU) to transparently duplicate and forward media streams between participants. Since an SFU does not process the media streams, so as to reduce delay, its latency is mainly determined by the underlying network I/O mechanism that moves packets to/from the network hardware and user space. Such mechanisms are usually based on POSIX sockets, which were not designed for high performance networking. We designed and implemented `pktswitch`, a minimal, socket-based SFU and measured its performance. We then modified `pktswitch` to employ `netmap`, a framework for fast packet I/O, to overcome the performance bottlenecks imposed by the socket-based design. The modified implementation handles packets in user space, with minimal kernel interaction. We describe and contrast the two implementations and then compare their performance in terms of packet processing overhead and delay. Our results show that the `netmap`-based implementation reduces packet processing overhead by 76% and delay by 89% compared to the socket-based implementation, thus allowing an SFU to host much higher loads (e.g. more users with more media streams and higher bit rates) without introducing delays.

**Index Terms**—Networked music performance, selective forwarding unit, latency.

## I. INTRODUCTION

In recent years, high performance network I/O has been a popular research topic [1]. When evaluating results, the vast majority of this research focuses on measuring isolated workloads, thus failing to relate high performance I/O research to real-world applications. As part of the MusiNet project [2], [3], we have instead focused on investigating the impact of such mechanisms to the performance of a *Networked Music Performance* (NMP) [4] system, that is, a system allowing the collaboration of musicians in real time. NMP systems are in essence ultra-low delay conferencing systems, where audio/video streams of varying quality have to be relayed between participants in-order and with minimal latency. In modern conferencing systems, a high performance machine, known in the literature as a *Selective Forwarding Unit* (SFU) [5], handles incoming streams and relays them to the participants with minimal processing. Unlike legacy systems, which performed audio mixing and video transcoding to produce a single outgoing stream for all participants, an SFU simply replicates and transmits incoming streams according to client preferences. Hence, SFUs are I/O intensive, making

them a potential candidate for deploying high-performance network I/O mechanisms.

Due to the critical role of SFUs in NMP latency, we have previously proposed replacing the traditional POSIX socket-based SFU implementations with alternatives based on high-performance I/O schemes [6], such as the Click modular router [7], the netFPGA boards [8] and the netmap scheme [9]. We have focused on netmap, designing a netmap-based SFU that offers the same functionality as a simple socket-based SFU [10]. In this paper we present two implementations of that minimal SFU, `pktswitch`, using either POSIX sockets or netmap, and examine their performance in terms of packet processing overhead and delay in a real Linux-based network testbed. Our results indicate that the netmap-based implementation reduces packet processing overhead by 76% and delay by 89%, thus allowing an SFU to host much higher loads without introducing delays.

The remainder of this paper is structured as follows. In Section II we provide a high-level description of network hardware and its driver interface, and contrast how network I/O is implemented when using sockets and netmap. In Section III we describe the basic functionality of `pktswitch` and describe our implementation of a baseline SFU implementation based on POSIX sockets. Then, we describe our netmap-based SFU implementation. We discuss the performance bottlenecks of the socket-based SFU and show how netmap can help overcome its limitations. In section IV we experimentally evaluate and compare the performance of the two implementations in terms of packet processing overhead and delay and examine how the system scales with increasing packet rates and clients. We conclude and discuss future work in Section V.

## II. DEMYSTIFYING NETWORK I/O

In this section, we briefly discuss the process of sending and receiving packets in a modern *Operating System* (OS), using Linux as a reference. Rather than explain specific implementation details, our goal is to understand the basic costs that dominate network I/O, that is, the process of moving packets from user space to the *Network Interface Card* (NIC), and vice-versa. From a high level perspective, this process can be broken down into three parts:

- 1) Moving packets between the application and the networking stack.

- 2) Processing within the networking stack.
- 3) Moving packets between the networking stack and the NIC.

We will begin by discussing the POSIX socket interface (1,2) and then we will discuss interfacing with NICs (3). Finally, we will discuss netmap, a high-performance network I/O framework, that streamlines this process.

#### A. POSIX Sockets

The default mechanism for network I/O provided by most systems is the POSIX socket interface. Using sockets, a system call is required for each packet received or sent by an application. Each system call triggers a kernel control path that handles the actual reception/transmission. We examine below each direction, focusing on UDP packets.

In order to send a UDP packet, the user application must issue a `sendto` system call. At that point, execution control is passed to the kernel and, among other things, the user arguments are validated. Next, an OS-specific packet buffer is allocated in kernel memory and the packet payload is copied into the packet buffer. Then, the packet buffer is passed to the OS networking stack and processed by the relevant protocols in all layers of the networking stack (UDP, IP, Ethernet). In case of a UDP packet, the UDP header is added to the packet and, optionally, a UDP checksum is computed. Next, the IP layer adds the IP header, performs IP routing to select an outgoing interface and computes the IP header checksum. Finally, the Ethernet layer performs ARP to determine the Ethernet destination address and adds the Ethernet header.

On the receive path, the user application must issue a `recvfrom` system call, passing again execution control to the kernel. Once the NIC driver moves the packet into host memory, it passes the packet to the OS networking stack. From there, the Ethernet layer checks the header and demultiplexes the packet based on its protocol, in our case IP. Next, the IP layer checks the IP header, performs IP routing and demultiplexes the packet to the UDP protocol handler. Then, the UDP layer checks the UDP header and copies the UDP payload to a user buffer destined for the appropriate application.

We can group these overheads into the following categories:

- **Kernel Crossings:** Switching from user to kernel mode, validating user arguments, and returning to user mode.
- **Memory Allocations:** Dynamically allocating a packet buffer upon packet reception/transmission.
- **Network Processing:** Copying, routing, checksumming, de-multiplexing packets, etc.
- **Packet Conversion:** Converting packets from their device specific format (see Section II-B) to the OS packet buffer format (`sk_buff` in Linux) and vice-versa.

#### B. Interfacing with NICs

In order for any packet to be transmitted over the physical link, it must be first moved from the host to the NIC. NICs communicate with the host computer over a bus, such as PCI. Data transfer is performed using a DMA engine in the hardware. Internally, NICs provide a fixed-size memory block

that is used to allocate receive (RX) and transmit (TX) FIFO circular queues (or rings), to temporarily store packets. On the host side, the NIC driver maintains RX/TX packet buffers and associated buffer descriptors containing various control information. Buffer descriptors are also organized as circular queues, forming RX/TX rings. The ownership of the descriptor rings (and thus their associated buffers) is split between the NIC and the driver, using head and tail ring indices. The driver owns all descriptors in the range  $[tail \dots head - 1]$ , while the NIC owns all descriptors in the range  $[head \dots tail - 1]$ . NICs use on-chip memory to fetch and write back RX/TX descriptors from the driver's RX/TX ring in host memory. Furthermore, a set of programmable registers is exported for each ring, storing the address of the ring in host memory, the length of the ring and the head and tail indices. On the receive path, the DMA engine fetches RX descriptors, transfers packets from the NIC RX ring to the host RX ring and updates the corresponding descriptors. On the transmit path, the DMA engine fetches TX descriptors, transfers packets from the host TX ring to the NIC TX ring and updates the TX descriptors.

#### C. An alternative solution: Netmap

Netmap is a framework for high performance I/O that moves packets from the NIC to user space and vice-versa, aiming to achieve very high throughput [9]. The idea behind netmap (and similar frameworks), is to move packet processing code from the kernel to user space, using OS primitives only for synchronization. From a high level perspective, netmap works by disconnecting the host networking stack from the packet data path; instead of delivering packets to the host networking stack, packets are delivered to the client application. For this purpose, a pair of RX/TX rings are exposed directly to user space. For cases where communication with the OS is needed, netmap exports an additional pair of RX/TX rings that can be used to pass packets to/from the host networking stack.

Netmap rings are similar to the driver RX/TX rings described a. Among other metadata, they include ring indices and pointers to packet buffers. The client application owns all buffers in the range  $[head \dots tail - 1]$ , while netmap owns buffers in the range  $[tail \dots head - 1]$ . To protect itself against misbehaving clients, the kernel is responsible for maintaining the integrity of the rings. For that purpose, the kernel maintains shadow rings and uses them to validate user requests.

Performance-wise, there are three distinct advantages to using netmap:

- Netmap does not use a `sk_buff` representation to store packets and their associated metadata, but rather a simple and NIC-friendly `netmap_ring` representation. This makes converting packets from their device specific representation to the `netmap_ring` format very efficient.
- Memory buffers are pre-allocated in linear memory, so no memory allocation/deallocation is needed when performing I/O.
- The direct exposure of RX/TX rings to client applications allows batch processing with minimum kernel mediation and the development of high performance specialized network stacks entirely in user space.

Using the netmap API is straightforward. The netmap client must first put the NIC in netmap mode and map the netmap device `/dev/netmap` into the process address space. The client can then start receiving/sending packets. Synchronization between the client and the NIC is implemented by using the standard `poll/select` system calls. The client and kernel rings can also be synchronized by using `ioctl`.

### III. IMPLEMENTING A NETWORKED MUSIC PERFORMANCE PLATFORM

In order to test our approach, we implemented a very simple conferencing platform, based on the needs of NMP. First, we developed a client application which is used by NMP participants to enter a conferencing session and stream/receive media. The client is split into three threads of execution. The main thread controls and configures client parameters (e.g. number of streams to send, packet transmission intervals, packet payload sizes) and outputs logging information (e.g. in/out Mbps, packet counters, packet loss). The generator thread is responsible for filling the output buffers with data and notifying the application when new data is ready for transmission. Finally, the engine thread does all the network I/O. Since our platform was created for testing purposes only, incoming packets are logged and immediately dropped, while outgoing packets have dummy media payload contents.

Second, we developed a SFU that handles incoming client streams. Much like the client, the SFU application is divided into two threads; the main thread is responsible for controlling and configuring the SFU parameters, while the engine thread is responsible for handling client packets and invoking the underlying network I/O mechanism as needed. In particular, upon receiving a media packet, the SFU must decide based on its origin, type and pre-configured client preferences, the subset of clients that the packet should be forwarded to. This allows NMP participants to select which media streams they want to receive. For example, they can omit video streams from some participants or, when layered coding is used, they can omit the video enhancement layer if they lack bandwidth. Finally, the SFU must replicate the packet as needed, and transmit it to the selected subset of clients.

The clients and SFU communicate by using control packets. Client identification is done by using a *Stream Source ID* (SSID). In order to route incoming media packets, the SFU is responsible for associating SSIDs with valid (IP address, port) pairs. The first step in participating in a media-streaming session, is to register with the SFU; that is, to receive a SSID. Once registration is complete, the client starts receiving other streams and can start transmitting. Each media packet is characterized by its payload type, which can be *Audio Standard Quality* (ASQ), *Video Standard Quality* (VSQ) and *Video High Quality* (VHQ). As its name implies, in scalable video coding, the VHQ stream acts as an enhancement layer for the VSQ layer. Our clients send all three packet types one after the other at user-configured, fixed intervals.

#### A. A socket-based approach

The POSIX socket interface provides a high level of abstraction and hides many details involved in receiving/sending

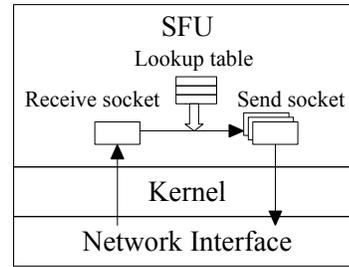


Fig. 1. A socket-based SFU.

packets. As a result, implementing a socket-based SFU is rather straightforward. When idle, we wait for incoming packets; since data can arrive at both the control and media ports we use `poll` to multiplex these inputs. When packets arrive, we move them into user space using `recvfrom`. Note that we can only move one packet at a time with each kernel crossing. Once the packet has been moved into user space, we must perform a lookup based on its SSRC and type to determine to which clients it should be sent to. We send the packet to every destination using repeated `sendto` calls. Again, note that we cannot send multiple packets at once. When the packet has been sent to all destinations, we can start processing the next packet; if none is available, we return to idle state. An overall view of the socket-based SFU is given in Fig. 1.

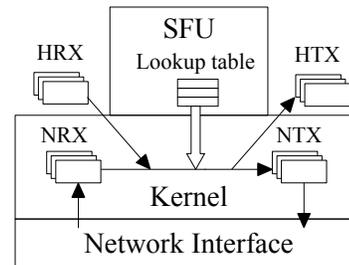


Fig. 2. A netmap-based SFU.

#### B. A netmap-based approach

Using netmap to implement the SFU is slightly more complicated: we must handle packets from multiple queues and route them to the appropriate destination. For the sake of simplicity, in this section we will assume that there is only one hardware queue, i.e., a single RX/TX ring pair per NIC. When idle, the SFU waits for packets from either the *netmap RX* (NRX) ring, which is connected to the NIC, or the *host RX* (HRX) ring, which is connected to the host networking stack. Assume that a media packet arrives; the SFU will return from `poll` and will determine the packet source, which in this case is the NRX ring. With netmap, when returning from a `poll` one or more packets can be moved to user space, without needing further kernel mediation. There are three possible destinations for packets coming from the NRX ring: the host OS (i.e., the HTX ring), the SFU control packet handler or the SFU media packet handler. Routing is done in

	Socket-based SFU	Netmap-based SFU
<b>Kernel Crossings</b>	$n$	1
<b>Memory Allocations</b>	$n$	Buffers are pre-allocated during startup
<b>Network Processing</b>	$n$ host networking stack iterations	minimal user space routing
<b>Packet Conversion</b>	complex <code>sk_buff</code> structure	simple NIC-friendly <code>netmap_ring</code> structure

TABLE I

SOCKET-BASED AND NETMAP-BASED SFU COSTS WHEN FORWARDING A MEDIA PACKET TO  $n$  DESTINATIONS.

user space by inspecting the packet headers and requires a few lookups in a table.

In the case of media packets, the packet is passed to the SFU media packet handler. From there, the handler will attempt to send the packet to all its destinations in one pass. First, the packet is copied  $n$  times, where  $n$  is the number of destinations, filling buffers in the NTX ring. Instead of routing each copy, one lookup is sufficient to update the packet's headers with the destination address. At the end of this process, the incoming packet buffer is also moved to the NTX ring and updated in order to be sent back to the origin client; this allows each client to estimate the end-to-end delay of its packets. If there are not enough NTX buffers to complete the process in one pass, control information is saved and netmap returns to polling, this time waiting for NTX buffers to become available.

Control packets are processed by the SFU control handler, updating its lookup table. Finally, packets destined to the host OS are copied to the HTX ring, i.e., the host networking stack, for processing by other applications. An overall view of the netmap-based SFU is given in Fig. 2.

### C. Performance Analysis

For most applications, the socket interface provides sufficient performance, as well as robustness and high portability. In some cases however, where packets arrive at a very high rate and little (or no) packet payload processing is required, sockets may become a performance bottleneck. In sections II-A and II-C we explained how the netmap scheme can avoid some of the overheads induced by sockets. This is particularly relevant for the SFU used in NMP applications, where no packet payload processing takes place, each received packet is copied multiple times and the media streams involved can have high bit rates, especially when video is involved.

Table I summarizes the basic differences between the two implementations, in the context of forwarding an incoming media packet in a scenario with  $n$  participants, using the overhead categories explained in section II-A. First, with sockets we need one system call per recipient, while with netmap we can make all outgoing packets available to the underlying NIC driver with a single system call. Second, with sockets we need to allocate one buffer per outgoing packet, while with netmap all buffers are pre-allocated during startup. Third, with sockets we need to pass each outgoing packet through the host networking stack, while with netmap all routing takes place at user space via a simple lookup operation. Fourth, with sockets we need to convert each `sk_buff` structure used by the host networking stack to the structure used by the NIC, while netmap uses a `netmap_ring` structure that is much more similar to the NIC structure.

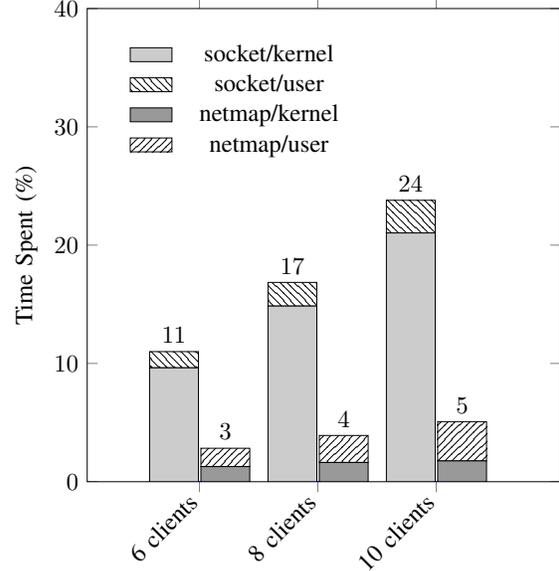


Fig. 3. CPU utilization with each client streaming at 750 pps (1.12 Mbps).

In order to understand how these overheads affect performance, we define packet processing *delay* as the time difference between receiving a packet and scheduling the last “echo” packet for transmission. Using the socket interface, the total processing delay for an incoming packet is dominated by system call execution overheads. Hence a rough approximation of the delay can be calculated by:

$$d_s(n) = k_{rcv} + n * k_{snd}$$

where  $k_{rcv}$  and  $k_{snd}$  are the respective system call delays and  $n$  is the number of participants. For example, for a conferencing session with 8 participants, given that the average system call overheads were measured to be  $k_{rcv} = 0.7 \mu s$  and  $k_{snd} = 2.3 \mu s$ , we expect a delay of at least  $19.1 \mu s$ . In contrast, with the netmap framework the total processing delay for an incoming packet is dominated by user space execution. A rough approximation of the cost can be calculated by:

$$d_n(n) = u_{rx} + n * u_{tx}$$

where  $u_{rx}$  is the cost of processing an incoming packet and  $u_{tx}$  is the cost of preparing an outgoing packet. For example, for a conferencing session with 8 participants, given that the average user space execution overheads were measured to be  $u_{rx} = 0.3 \mu s$  and  $u_{tx} = 0.2 \mu s$ , we expect a delay of at least  $1.9 \mu s$ .

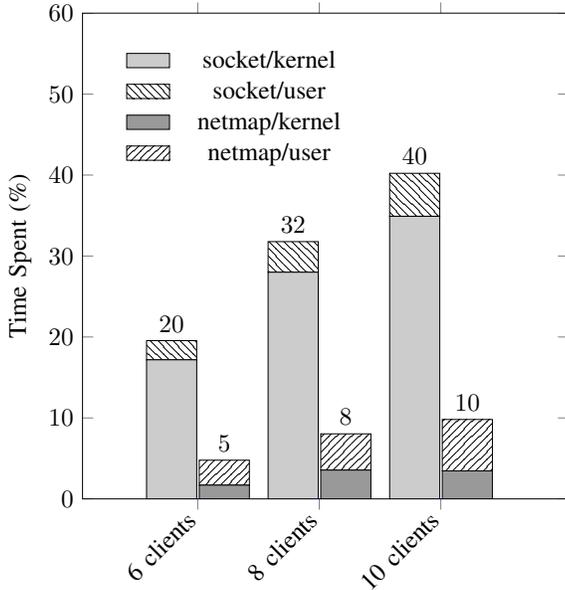


Fig. 4. CPU utilization with each client streaming at 1500 pps (2.24 Mbps).

#### IV. EXPERIMENTAL EVALUATION

Using our prototype implementations, we experimentally evaluated their performance in terms of packet processing overhead and delay. To assess processing overhead we measured the CPU utilization, broken down in its kernel and user parts, in an otherwise idle system. This metric indicates how much we can load the system before the CPU becomes a bottleneck, at which point packets will have to wait to be processed due to CPU limitations. To assess delay we measured the difference between the time a media packet was picked up to be processed and the time its final “echo” packet was scheduled to be sent. This metric indicates how much time the SFU itself takes to process the packets, ignoring the NIC delays which are the same in both implementations.

Our testbed consisted of 11 identical machines using the Realtek RTL8111/8168B PCI Express Gigabit Ethernet controller and running Linux kernel 3.2.63. Our test machines were connected in an isolated Gigabit LAN. One machine acted as the SFU while the rest acted as the clients. In our experiments we varied two factors: the number of clients participating in each experiment and the bit rate of the media transmitted by each client. Each experiment consisted of a warmup period, followed by a 60 s period of measurements.

Since the absolute numbers we are measuring are at the microsecond scale, we tried our best to ensure that performance monitoring had a negligible (if any) effect on results. For efficient and high-resolution delay measurements, we counted CPU cycles, whereas for system-wide measurements we utilized the `/proc` pseudo-file system exported by Linux. To ensure repeatability, we ran each experiment multiple times, collected data over multiple runs and verified that results had very low variance.

#### A. Processing Overhead

We measured packet processing overhead as the fraction of time (%) our CPU spent processing packets. Assuming that the NIC is able to cope with traffic at the hardware level, the less CPU time an SFU requires to operate, the more clients it can support, as well as higher bit rates. To further highlight how the two implementations differ, we present kernel time separately from user time. Kernel time, the time spent processing packets in kernel-mode, captures the costs of executing system calls. In the socket-based SFU, that includes `poll`, `recvfrom` and `sendto`, whereas in the case of the netmap-based SFU, it only includes `poll` and `ioctl`. Conversely, user time measures the time spent processing in user-mode. In the socket-based SFU, user-time captures the cost of our SFU code, including application-level routing. In the netmap-based SFU, user-time also includes the cost of SFU packet routing (both RX and TX) and packet copying.

Figure 3 shows CPU utilization with 6, 8 and 10 clients, when clients send media packets at 4 ms intervals and the media payload sizes are set to 80, 160 and 320 bytes for the ASQ, VSQ and VHQ streams, respectively. This means that each client sends 250 packets per second (pps) for each stream, or 750 pps in total, for a combined data rate of 1.12 Mbps per client<sup>1</sup>. In all cases, the netmap-based SFU requires significantly less CPU time to keep up with the traffic, using on average only 23.4% of the CPU time that the socket-based SFU requires. As expected, in the socket-based SFU, most of the time is spent running kernel code inside system calls, with only a fraction spent executing in user space. In contrast, the netmap-based SFU spent most of its execution time running user space code, where packet replication and routing take place. Figure 4 shows the results of the same experiment, but with clients sending media packets at 2 ms intervals, effectively doubling the data rate to 1500 pps per client, or 2.24 Mbps. The relative performance of the two implementations is roughly the same, with the netmap-based SFU using on average only 24.7% of the CPU time that the socket-based SFU requires.

As these experiments clearly show, while CPU utilization increases with the number of clients and their bit rate with both implementations, the netmap-based SFU should be able to handle around four times the load of the socket-based SFU, whether in terms of clients or bit rates, before the CPU is saturated, leading to CPU induced packet delays.

#### B. Delay

Packet processing delay measures the time needed for a media packet to be forwarded to all its destinations. For the socket-based SFU this means queuing the `n sk_buff` structures to the OS networking stack for transmission. While the number of `sk_buffs` the OS allocates for the SFU is limited (after that, it may start dropping packets), it is more than enough to provide temporary buffering when the NIC TX ring is congested. Hence, the observed processing delay of the socket-based SFU is not affected by such congestion

<sup>1</sup>This figure excludes header overheads, which, due to the small size of the packets, are not negligible.

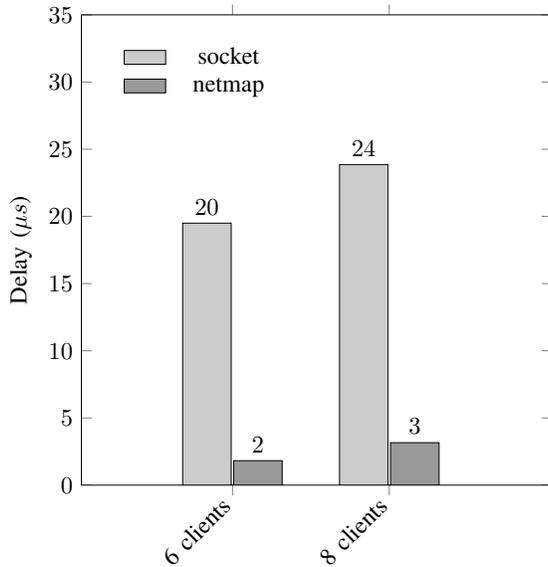


Fig. 5. Packet processing delay each client streaming at 750 pps (1.12 Mbps).

events, although the delay does affect the actual packets. On the other hand, when the NIC TX rings become congested, netmap blocks transmission until TX ring buffers become available. Hence, this cost is included in the observed delays of the netmap-based SFU. For this reason, we restrained from reporting latency results from experiments where the TX rings were heavily congested, as the comparison was heavily biased in favor of the socket-based SFU and our results, although still showing improved performance with the netmap-based SFU, exhibited high variance.

Figure 5 shows the packet processing delay measured for sessions with 6 and 8 clients, transmitting media packets at 4 ms intervals, or 750 pps per client. Again, the netmap-based SFU shows significant improvements, with a delay that is on average only 11.3% of the delay of the socket-based SFU. Based on our discussion at Section III-C we can expect similar speedups as the number of clients increases. Furthermore, although not presented in this paper, in our experiments we found that as long as the NIC TX queues were not congested and the CPU usage was reasonable, packet processing delays were not affected.

## V. CONCLUSIONS

We presented the design and implementation of a socket-based SFU for NMP applications, as well as an alternative SFU implementation using netmap, a high-performance network I/O scheme. We highlighted the bottlenecks of using sockets for the SFU and explained how low-level network I/O can overcome them. We simulated a number of real-world conferencing sessions, and measured the packet processing overhead and delay of both implementations against the number of participating clients and the bit rates emitted by the clients. Our measurements show that the netmap-based implementation is significantly lighter and faster than its socket-based counterpart, reducing packet processing overhead by 76% and delay by 89%. These results indicate that high performance

I/O mechanisms have great potential for improving the performance and scalability of SFUs built for NMP applications.

## ACKNOWLEDGMENT

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS - University of Crete - MUSINET.

## REFERENCES

- [1] S. Peter, J. Li, I. Zhang, D. R. K. Ports, D. Woos, A. Krishnamurthy, T. Anderson, and T. Roscoe, “Arrakis: The operating system is the control plane,” in *Proc. of the USENIX Conference on Operating Systems Design and Implementation (OSDI)*, 2014.
- [2] The MusiNet project. [Online]. Available: <http://musinet.aueb.gr/>
- [3] D. Akoumianakis, C. Alexandraki, V. Alexiou, C. Anagnostopoulou, A. Eleftheriadis, V. Lalioti, A. Mouchtaris, D. Pavlidi, G. C. Polyzos, P. Tsakalides, G. Xylomenos, and P. Zervas, “The MusiNet project: Towards unraveling the full potential of networked music performance systems,” in *Proc. of the International Conference on Information, Intelligence, Systems and Applications (IISA)*, 2014.
- [4] A. Renaud, A. Carôt, and P. Rebelo, “Networked music performance: State of the art,” in *Proc. of the AES Conference on Intelligent Audio Environments*, 2007.
- [5] A. Eleftheriadis, R. M. Civanlar, and O. Shapiro, “Multipoint videoconferencing with scalable video coding,” *Journal of Shejiang University SCIENCE A*, vol. 7, pp. 696–705, 2006.
- [6] G. Xylomenos, C. Tsilopoulos, Y. Thomas, and G. C. Polyzos, “Reduced switching delay for networked music performance,” in *Packet Video Workshop (Poster Session)*, 2013.
- [7] E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek, “The Click modular router,” *ACM Transactions on Computer Systems*, vol. 18, pp. 263–297, 2000.
- [8] J. Lockwood, N. McKeown, G. Watson, G. Gibb, P. Hartke, J. Naoos, R. Raghuraman, and J. Luo, “NetFPGA—an open platform for gigabit-rate network switching and routing,” in *Proc. of the IEEE Microelectronic Systems Education Conference (MSE)*, 2007.
- [9] L. Rizzo, “Netmap: a novel framework for fast packet I/O,” in *Proc. of the USENIX Annual Technical Conference (ATC)*, 2012.
- [10] G. Baltas and G. Xylomenos, “Ultra low delay switching for networked music performance,” in *Proc. of the International Conference on Information, Intelligence, Systems and Applications (IISA)*, 2014.