THE GREAT ESCAPE

A Case Study Of VM Escape & EoP Vulnerabilities

Billy
Security Research
@st424204

Muhd. Ramdhan
Security Research
@n0psledbyte

About Us

Billy

- Vulnerability researcher from STAR Labs,
- Interested at hypervisor and kernel security
- CTF player from Balsn

Ramdhan

- Vulnerability researcher from STAR Labs
- Interested at hypervisor and kernel security
- CTF player from Super Guesser

STAR
LABS

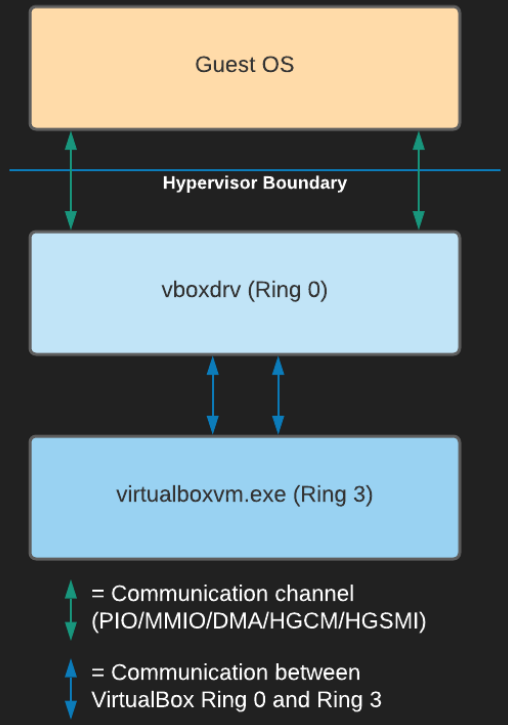
Outline

- Oracle Virtualbox Overview
- Attack Surface
- CVE-2021-2321: OOB Read Information Disclosure Vulnerability
- CVE-2021-2250: VirtualBox SLIRP Heap-based Overflow
- Linux Kernel Privilege Escalation

Oracle Virtualbox Overview

- One of the popular virtualization software for desktop
- C/C++, Free, Open source (It's good for code auditing and find some memory corruption bugs ;))
- VirtualBox Extensions shipped as binaries, to support USB 2.0 and 3.0 devices, VirtualBox RDP, etc.
- Virtualbox is using Hardware-assisted Virtualization based on Intel VT-X and AMD-V

Oracle Virtualbox Architecture Overview



- For performance reason it don’t switch to R3 directly
- Ring 0 mostly handling VT-X/AMD-V code and some small amount of code that handling I/O interaction from guest
- The bigger amount of code run in Ring 3
- The code that handling request from some communication channel should be interesting for attacker

Attack Surface

- Emulated devices is the most interesting attack surface, based on vulnerability in the past
- Virtualbox have a lot of emulated devices, some of them is not enabled by default:
  - Networking : E1000, virtio-net, SLiRP
  - Audio : Intel HDA, AC97
  - Graphic : VGA Device
  - USB : OHCI, EHCI, xHCI
  - Storage : AHCI
- There are other interesting attack surface through such as HGCM (Host Guest Communication Manager) to interact with specific virtualbox service such as Shared Folder, Shared Clipboard, etc

Bug Hunting

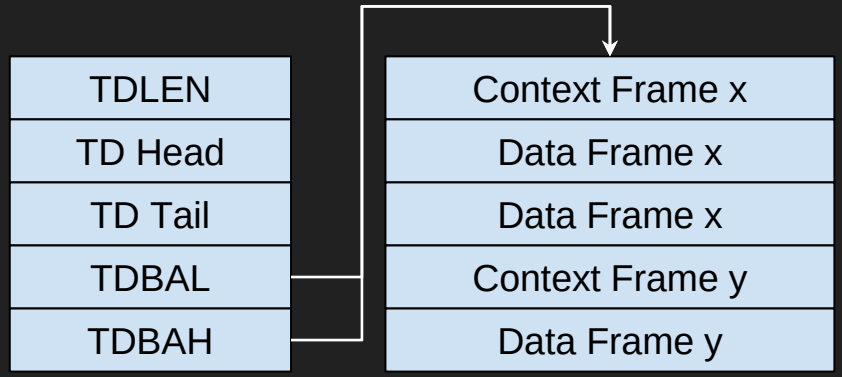
- Attack surface in network emulated devices always used in Pwn2Own 2019, 2020
- So we choose network devices first for hunting VM escape bug
- We found two bugs by code auditing that can be used for VM escape
  - CVE-2021-2321: Oracle VirtualBox e1000 Out-Of-Bounds Read Information Disclosure Vulnerability
  - CVE-2021-2250: Oracle VirtualBox SLiRP Networking Heap-based Overflow Privilege Escalation Vulnerability

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- Bug resides in code that handling e1000 frame
- Found by code auditing
- Inspired by CVE-2020-2894: e1000 Out-Of-Bounds Read Vulnerability (Shout out to Pham Hong Phi!)
- Have some (quite) complex logic bug, that can be turned into Out-Of-Bounds Read Vulnerability

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

E1000 Basic



- One packet can contain one context frame followed by multiple data frames
- Last data frame in packet have End Of Packet flag

```
struct E1kTDData
{
    uint64_t u64BufAddr;                        /**< Address of data buffer */
    struct TDDCmd_st
    {
        unsigned u20DTALEN : 20; /** The total length of data pointed to by this
descriptor. */
        unsigned u4DTYP : 4; /** The descriptor type - E1K_DTYP_DATA (1). */
        unsigned fEOP : 1; /** End of packet.  Note TSCTFC update.  */
        unsigned fIFCS : 1; /** Insert Ethernet FCS/CRC (requires fEOP to be set). */
        unsigned fTSE : 1; /** Use the TSE context when set and the normal when clear.
*/
        unsigned fRS : 1; /** Report status (dw3.STA). */
        unsigned fRPS : 1; /** Reserved. 82544GC/EI defines this report packet set
(RPS).  */
        unsigned fDEXT : 1; /** Descriptor extension, must be set for this descriptor
type. */
        /** VLAN enable, requires CTRL.VME, auto enables FCS/CRC.
         *  Insert dw3.SPECIAL after ethernet header. */
        unsigned fVLE : 1;
        unsigned fIDE : 1; /** Interrupt delay enable. */
    } cmd;
    struct TDDDw3_st
    {
        unsigned fDD : 1;                       /**< Descriptor done. */
        unsigned fEC : 1;                      /**< Excess collision. */
        unsigned fLC : 1;                        /**< Late collision. */
        unsigned fTURSV : 1; /** Reserved, except for the usual oddball (82544GC/EI)
where it's called TU. */
        unsigned u4RSV : 4;                  /**< Reserved field, MBZ. */
        unsigned fIXSM : 1;                   /**< Insert IP checksum. */
        unsigned fTXSM : 1;                /**< Insert TCP/UDP checksum. */
        unsigned u6RSV : 6;                      /**< Reserved, MBZ. */
        unsigned u16Special : 16;   /**< VLAN: Id, Canonical form, Priority. */
    } dw3;
};
```

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- Fetching valid frame

```
while (e1kLocateTxPacket(pThis))
{
    fIncomplete = false;
    /* Found a complete packet, allocate it. */
    rc = e1kXmitAllocBuf(pThis, pThisCC, pThis->fGSO);
    /* If we're out of bandwidth we'll come back later. */
    if (RT_FAILURE(rc))
        goto out;
    /* Copy the packet to allocated buffer and send it. */
    rc = e1kXmitPacket(pDevIns, pThis, fOnWorkerThread, &txdc);
    /* If we're out of bandwidth we'll come back later. */
    if (RT_FAILURE(rc))
        goto out;
}
```

```
static bool e1kLocateTxPacket(PE1KSTATE pThis)
{
    ...
    for (int i = pThis->iTxDCurrent; i < pThis->nTxDFetched; ++i)
    {
        E1KTXDESC *pDesc = &pThis->aTxDescriptors[i];
        switch (e1kGetDescType(pDesc))
        {
            case E1K_DTYP_CONTEXT:
                ...
            case E1K_DTYP_LEGACY:
                ...
            case E1K_DTYP_DATA:
                ...
            default:
                AssertMsgFailed(("Impossible descriptor type!"));
                continue;
        }
        if (pDesc->legacy.cmd.fEOP)
        {
            ...
            return true;
        }
    }
}
```

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- `e1kXmitPacket` will processing one packet frame

```
while (e1kLocateTxPacket(pThis))
{
    fIncomplete = false;
    /* Found a complete packet, allocate it. */
    rc = e1kXmitAllocBuf(pThis, pThisCC, pThis->fGSO);
    /* If we're out of bandwidth we'll come back later. */
    if (RT_FAILURE(rc))
        goto out;
    /* Copy the packet to allocated buffer and send it. */
    rc = e1kXmitPacket(pDevIns, pThis, fOnWorkerThread, &txdc);
    /* If we're out of bandwidth we'll come back later. */
    if (RT_FAILURE(rc))
        goto out;
}
```

```
static int e1kXmitPacket(PPDMDEVINS pDevIns, PE1KSTATE pThis, bool fOnWorkerThread, PE1KTXDC pTxdc)
{
    PE1KSTATECC pThisCC = PDMDEVINS_2_DATA_CC(pDevIns, PE1KSTATECC);
    int rc = VINF_SUCCESS;

        /* iterate frame */
    while (pThis->iTxDCurrent < pThis->nTxDFetched)
    {
        E1KTXDESC *pDesc = &pThis->aTxDescriptors[pThis->iTxDCurrent];
        E1kLog3(("%s About to process new TX descriptor at %08x%08x, TDLEN=%08x, TDH=%08x, TDT=%08x\n",
                 pThis->szPrf, TDBAH, TDBAL + pTxdc->tdh * sizeof(E1KTXDESC), pTxdc->tdlen, pTxdc->tdh, pTxdc->
>tdt));
                /* processing frame */
        rc = e1kXmitDesc(pDevIns, pThis, pThisCC, pDesc, e1kDescAddr(TDBAH, TDBAL, pTxdc->tdh), fOnWorkerThread);
                ...
        ++pThis->iTxDCurrent;
        if (e1kGetDescType(pDesc) != E1K_DTYP_CONTEXT && pDesc->legacy.cmd.fEOP)
            break;
    }

    return rc;
}
```

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- Useless logic error? no

```
e1kXmitPacket -> e1kXmitDesc ->
-> e1kFallbackAddToFrame -> e1kFallbackAddSegment
```

```
#define E1K_MAX_TX_PKT_SIZE    0x3fa0
...
/** TX: Transmit packet buffer use for TSE fallback and loopback. */
uint8_t     aTxPacketFallback[E1K_MAX_TX_PKT_SIZE];
```

```
static int e1kFallbackAddSegment(PPDMDEVINS pDevIns, PE1KSTATE pThis, RTGCPHYS PhysAddr, uint16_t u16Len, bool
fSend, bool fOnWorkerThread)
{
    int rc = VINF_SUCCESS;
    PE1KSTATECC pThisCC = PDMDEVINS_2_DATA_CC(pDevIns, PE1KSTATECC);
    /* TCP header being transmitted */
    struct E1kTcpHeader *pTcpHdr = (struct E1kTcpHeader *)(pThis->aTxPacketFallback + pThis->contextTSE.tu.u8CSS);
    /* IP header being transmitted */
    struct E1kIpHeader *pIpHdr = (struct E1kIpHeader *)(pThis->aTxPacketFallback + pThis->contextTSE.ip.u8CSS);

    AssertReturn(pThis->u32PayRemain + pThis->u16HdrRemain > 0, VINF_SUCCESS);

    if (pThis->u16TxPktLen + u16Len <= sizeof(pThis->aTxPacketFallback)) // [1]
        PDMDevHlpPhysRead(pDevIns, PhysAddr, pThis->aTxPacketFallback + pThis->u16TxPktLen, u16Len);
    pThis->u16TxPktLen += u16Len; // [2]
    ...
}
```


STAR LABS

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- We don't have buffer overflow
  - `bound check in physicRead`
- But, we have OOB at `e1kInsertChecksum`
- How to pass large `u16Len` ?

```
static int e1kFallbackAddSegment(PPDMDEVINS pDevIns, PE1KSTATE pThis, RTGCPHYS PhysAddr,
uint16_t u16Len, bool fSend, bool fOnWorkerThread)
{
    int rc = VINF_SUCCESS;
    PE1KSTATECC pThisCC = PDMDEVINS_2_DATA_CC(pDevIns, PE1KSTATECC);
    /* TCP header being transmitted */
    struct E1kTcpHeader *pTcpHdr = (struct E1kTcpHeader *)(pThis->aTxPacketFallback +
pThis->contextTSE.tu.u8CSS);
    /* IP header being transmitted */
    struct E1kIpHeader *pIpHdr = (struct E1kIpHeader *)(pThis->aTxPacketFallback + pThis-
>contextTSE.ip.u8CSS);

    AssertReturn(pThis->u32PayRemain + pThis->u16HdrRemain > 0, VINF_SUCCESS);

    if (pThis->u16TxPktLen + u16Len <= sizeof(pThis->aTxPacketFallback))
        PDMDevHlpPhysRead(pDevIns, PhysAddr, pThis->aTxPacketFallback + pThis-
>u16TxPktLen, u16Len);
    pThis->u16TxPktLen += u16Len;
    ...

    if (fSend)
    {
        ...
        e1kInsertChecksum(pThis, pThis->aTxPacketFallback, pThis->u16TxPktLen,
                pThis->contextTSE.ip.u8CSO,
                pThis->contextTSE.ip.u8CSS,
                pThis->contextTSE.ip.u16CSE); // [2]
        ...
        e1kInsertChecksum(pThis, pThis->aTxPacketFallback, pThis->u16TxPktLen,
                pThis->contextTSE.tu.u8CSO,
                pThis->contextTSE.tu.u8CSS,
                pThis->contextTSE.tu.u16CSE); // [3]
    }
}
```

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- How to pass large `u16Len` ?

```
static int e1kFallbackAddToFrame(PPDMDEVINS pDevIns, PE1KSTATE pThis, E1KTXDESC *pDesc, bool fOnWorkerThread)
{
    uint16_t u16MaxPktLen = pThis->contextTSE.dw3.u8HDRLEN + pThis->contextTSE.dw3.u16MSS;
    int rc = VINF_SUCCESS;
    do
    {
        /* Calculate how many bytes we have left in this TCP segment */
        uint16_t cb = u16MaxPktLen - pThis->u16TxPktLen;
        if (cb > pDesc->data.cmd.u20DTALEN)
        {
            /* This descriptor fits completely into current segment */
            cb = (uint16_t)pDesc->data.cmd.u20DTALEN; /* u20DTALEN at this point is guarantied to fit into 16 bits. */
            rc = e1kFallbackAddSegment(pDevIns, pThis, pDesc->data.u64BufAddr, cb, pDesc->data.cmd.fEOP /*fSend*/, fOnWorkerThread); // [1]
        }
        else
        {
            rc = e1kFallbackAddSegment(pDevIns, pThis, pDesc->data.u64BufAddr, cb, true /*fSend*/, fOnWorkerThread); // [2]
            /*
             * Rewind the packet tail pointer to the beginning of payload,
             * so we continue writing right beyond the header.
             */
            pThis->u16TxPktLen = pThis->contextTSE.dw3.u8HDRLEN;
        }

        pDesc->data.u64BufAddr    += cb;
        pDesc->data.cmd.u20DTALEN -= cb;
    } while (pDesc->data.cmd.u20DTALEN > 0 && RT_SUCCESS(rc));

    ...
    return VINF_SUCCESS; /// @todo consider rc;
```

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- How to pass large `u16Len` ?

We need pass cb more than 0x3fa0 to get OOB

```
static int e1kFallbackAddToFrame(PPDMDEVINS pDevIns, PE1KSTATE pThis, E1KTXDESC *pDesc, bool fOnWorkerThread)
{
    uint16_t u16MaxPktLen = pThis->contextTSE.dw3.u8HDRLEN + pThis->contextTSE.dw3.u16MSS;
    int rc = VINF_SUCCESS;
    do
    {
        /* Calculate how many bytes we have left in this TCP segment */
        uint16_t cb = u16MaxPktLen - pThis->u16TxPktLen;
        if (cb > pDesc->data.cmd.u20DTALEN)
        {
            /* This descriptor fits completely into current segment */
            cb = (uint16_t)pDesc->data.cmd.u20DTALEN; /* u20DTALEN at this point is guarantied to fit into 16 bits. */
            rc = e1kFallbackAddSegment(pDevIns, pThis, pDesc->data.u64BufAddr, cb, pDesc->data.cmd.fEOP /*fSend*/, fOnWorkerThread); // [1]
        }
        else
        {
            rc = e1kFallbackAddSegment(pDevIns, pThis, pDesc->data.u64BufAddr, cb, true /*fSend*/, fOnWorkerThread); // [2]
            /*
             * Rewind the packet tail pointer to the beginning of payload,
             * so we continue writing right beyond the header.
             */
            pThis->u16TxPktLen = pThis->contextTSE.dw3.u8HDRLEN;
        }

        pDesc->data.u64BufAddr    += cb;
        pDesc->data.cmd.u20DTALEN -= cb;
    } while (pDesc->data.cmd.u20DTALEN > 0 && RT_SUCCESS(rc));

        ...
    return VINF_SUCCESS; /// @todo consider rc;
}
```

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- How to pass large `u16Len` ?

```
static int e1kFallbackAddToFrame(PPDMDEVINS pDevIns, PE1KSTATE pThis, E1KTXDESC *pDesc, bool fOnWorkerThread)
{
    uint16_t u16MaxPktLen = pThis->contextTSE.dw3.u8HDRLEN + pThis->contextTSE.dw3.u16MSS;
    int rc = VINF_SUCCESS;
    do
    {
        /* Calculate how many bytes we have left in this TCP segment */
        uint16_t cb = u16MaxPktLen - pThis->u16TxPktLen;
        if (cb > pDesc->data.cmd.u20DTALEN)
        {
            /* This descriptor fits completely into current segment */
            cb = (uint16_t)pDesc->data.cmd.u20DTALEN; /* u20DTALEN at this point is guarantied to fit into 16 bits. */
            rc = e1kFallbackAddSegment(pDevIns, pThis, pDesc->data.u64BufAddr, cb, pDesc->data.cmd.fEOP /*fSend*/, fOnWorkerThread); // l
        }
        else
        {
            rc = e1kFallbackAddSegment(pDevIns, pThis, pDesc->data.u64BufAddr, cb, true /*fSend*/, fOnWorkerThread); // [2]
            /*
             * Rewind the packet tail pointer to the beginning of payload,
             * so we continue writing right beyond the header.
             */
            pThis->u16TxPktLen = pThis->contextTSE.dw3.u8HDRLEN;
        }

        pDesc->data.u64BufAddr    += cb;
        pDesc->data.cmd.u20DTALEN -= cb;
    } while (pDesc->data.cmd.u20DTALEN > 0 && RT_SUCCESS(rc));

    VINF_                                    der rc;
```

We need pass cb more than 0x3fa0 to get OOB

We can control `u16MaxPktLen` but it never more than 0x3fa0

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- How to pass large `u16Len` ? another problem ..

```
static int e1kFallbackAddToFrame(PPDMDEVINS pDevIns, PE1KSTATE pThis, E1KTXDESC *pDesc, bool fOnWorkerThread)
{
    uint16_t u16MaxPktLen = pThis->contextTSE.dw3.u8HDRLEN + pThis->contextTSE.dw3.u16MSS;
    int rc = VINF_SUCCESS;
    do
    {
        /* Calculate how many bytes we have left in this TCP segment */
        uint16_t cb = u16MaxPktLen - pThis->u16TxPktLen;
        if (cb > pDesc->data.cmd.u20DTALEN)
        {
            /* This descriptor fits completely into current segment */
            cb = (uint16_t)pDesc->data.cmd.u20DTALEN; /* u20DTALEN at this point is guarantied to fit into 16 bits. */
            rc = e1kFallbackAddSegment(pDevIns, pThis, pDesc->data.u64BufAddr, cb, pDesc->data.cmd.fEOP /*fSend*/, fOnWorkerThread); // l
        }
        else
        {
            rc = e1kFallbackAddSegment(pDevIns, pThis, pDesc->data.u64BufAddr, cb, true /*fSend*/, fOnWorkerThread); // [2]
            /*
             * Rewind the packet tail pointer to the beginning of payload,
             * so we continue writing right beyond the header.
             */
            pThis->u16TxPktLen = pThis->contextTSE.dw3.u8HDRLEN;
        }

        pDesc->data.u64BufAddr    += cb;
        pDesc->data.cmd.u20DTALEN -= cb;
    } while (pDesc->data.cmd.u20DTALEN > 0 && RT_SUCCESS(rc));

    VINF_                                    der rc;
```

We need pass cb more than 0x3fa0 to get OOB

We can control `u16MaxPktLen` but it never more than 0x3fa0

Somehow, we need to make `pThis->u16TxPktLen` to create **int overflow**, so we can pass large `cb`

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- Another problem is to make `pThis->u16PktLen` bigger than `u16MaxPktLen` (`pThis->contextTSE.dw3.u16MSS + pThis->contextTSE.dw3.u8HDRLEN`)
- While processing one packet frame, there's no way to make `pThis->u16PktLen` bigger than `u16MaxPktLen`
- We already know, we can control `u16MaxPktLen` but it can't have more than `E1K_MAX_TX_PKT_SIZE` (0x3fa0)
- Somehow, we control `pThis->u16PktLen` to some value, and then in the next packet frame we control `u16MaxPktLen` to be less than `pThis->u16PktLen`

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- Can we control `pThis->u16PktLen` for the next processing packet frame?
- Last data frame contain `fEOP` enabled, and seems it will always clear `pThis->u16PktLen`, is there a way to make `pThis->u16PktLen` still alive for the next frame?

```
if (pDesc->data.cmd.u20DTALEN == 0 || pDesc->data.u64Buf
{
    E1kLog2(("% Empty data descriptor, skipped.\n", pThi
    if (pDesc->data.cmd.fEOP)
    {
        e1kTransmitFrame(pDevIns, pThis, pThisCC, fOnWor
        pThis->u16TxPktLen = 0;
    }
}
```

```
if (pDesc->data.cmd.fEOP)
{
    /* End of packet, next segment will contain header. */
    if (pThis->u32PayRemain != 0)
        E1K_INC_CNT32(TSCTFC);
    pThis->u16TxPktLen = 0;
    e1kXmitFreeBuf(pThis, PDMDEVINS_
```

```
else if (pDesc->legacy.cmd.fEOP)
{
    e1kXmitFreeBuf(pThis, pThisCC);
    pThis->u16TxPktLen = 0;
}
```

```
if (pDesc->data.cmd.fEOP)
{
    ...
    pThis->u16TxPktLen = 0;
```

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- We can just set `fDD` enabled in last data frame to avoid those checks!

```
static int e1kXmitDesc(PPDMDEVINS pDevIns, PE1KSTATE pThis, PE1KSTATECC pThisCC,
E1KTXDESC *pDesc,
        RTGCPHYS addr, bool fOnWorkerThread)
{
    int rc = VINF_SUCCESS;

    e1kPrintTDesc(pThis, pDesc, "vvv");

    if (pDesc->legacy.dw3.fDD)
    {
        E1kLog(("%s e1kXmitDesc: skipping bad descriptor ^^^\n", pThis->szPrf));
        e1kDescReport(pDevIns, pThis, pDesc, addr);
        return VINF_SUCCESS;
    }
    ...
    /* some check will happen to clear pThis->u16TxPktLen
     * if pDesc is last data frame */
}
```

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

Recap

- Logic error in mishandling `fDD` flag that allowed us to make int overflow at `e1kFallbackAddToFrame`
- Using int overflow, we can pass large value to `e1kFallbackAddSegment`
- Because of some missing check handling in `e1kFallbackAddSegment`, we can make `pThis->u16PktLen` large than its buffer, and it allowed us to create OOB Read in `e1kInsertCheckSum`

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

- We can pass pass large `u16Len` to make OOB at **[1]** and **[2]**

```
static int e1kFallbackAddSegment(PPDMDEVINS pDevIns, PE1KSTATE pThis, RTGCPHYS PhysAddr, uint16_t u16Len, bool fSend,
bool fOnWorkerThread)
{
  ...

  pThis->u16TxPktLen += u16Len;
  ...

  if (fSend)
  {
    ...
    pIpHdr->chksum = 0;
    e1kInsertChecksum(pThis, pThis->aTxPacketFallback, pThis->u16TxPktLen,
            pThis->contextTSE.ip.u8CSO,
            pThis->contextTSE.ip.u8CSS,
            pThis->contextTSE.ip.u16CSE); // [1]
    ...
    e1kInsertChecksum(pThis, pThis->aTxPacketFallback, pThis->u16TxPktLen,
            pThis->contextTSE.tu.u8CSO,
            pThis->contextTSE.tu.u8CSS,
            pThis->contextTSE.tu.u16CSE); // [2]

    // send packet to localhost (to retreive the checksum information)
  }
}
```

```
static void e1kInsertChecksum(PE1KSTATE pThis, uint8_t *pPkt, uint16_t u16PktLen, uint8_t
cso, uint8_t css, uint16_t cse, bool fUdp = false)
{
    ...
    int16_t u16ChkSum = e1kCSum16(pPkt + css, cse - css + 1);
    if (fUdp && u16ChkSum == 0)
        u16ChkSum = ~u16ChkSum;     /* 0 means no checksum computed in case of UDP (see
@bugref{9883}) */
    E1kLog2(("%s Inserting csum: %04X at %02X, old value: %04X\n", pThis->szPrf,
                u16ChkSum, cso, *(uint16_t*)(pPkt + cso)));
    *(uint16_t*)(pPkt + cso) = u16ChkSum;
}
```

CVE-2021-2321: OOB Read Information Disclosure Vulnerability

Recap (next)

- `e1kInsertCheckSum` can calculate checksum from data out of the bound from its buffer
- Using checksum value information we can leak two bytes behind the buffer at a time by calculating the difference between two checksums
- We can retrieve VBoxDD base address to bypass ASLR and building payload for our ROP Gadgets

CVE-2021-2250: VirtualBox SLIRP Heap-based Overflow

- SLIRP is one of the attack surface enabled by default in virtualbox
- Used for user-mode networking by emulating TCP/IP protocol
- This bug resides in emulating of ICMP protocol when the guest try to send ICMP request
- This bug only affected windows host only

CVE-2021-2250: VirtualBox SLIRP Heap-based Overflow

- Size of `struct ip` is only 20 bytes
- We can control `hlen` (IP header length) up to 60 bytes, so we can overwrite `pong->bufsize`
- `pong->bufsize` will be used as reply size, by overwriting it we can receive ICMP reply buffer larger than its size (heap overflow)

```
struct pong {
    PNATState pData;
    TAILQ_ENTRY(pong) queue_entry;
    struct ip reqiph;
    struct icmp_echo reqicmph;
    size_t bufsize;
    uint8_t buf[1];
};
```

```
void
icmpwin_ping(PNATState pData, struct mbuf *m, int hlen)
{
    struct ip *ip = mtod(m, struct ip *);
        ...

    reqsize = ip->ip_len - hlen - sizeof(struct icmp_echo);

    bufsize = sizeof(ICMP_ECHO_REPLY);
    if (reqsize < sizeof(IO_STATUS_BLOCK) + sizeof(struct icmp_echo))
        bufsize += sizeof(IO_STATUS_BLOCK) + sizeof(struct icmp_echo);
    else
        bufsize += reqsize;
    bufsize += 16; /* whatever that is; empirically at least XP needs it
*/

    pongsize = RT_UOFFSETOF(struct pong, buf) + bufsize;

    pong = RTMemAlloc(pongsize);
    if (RT_UNLIKELY(pong == NULL))
        return;

    pong->pData = pData;
    pong->bufsize = bufsize;
    m_copydata(m, 0, hlen, (caddr_t)&pong->reqiph); // [1]
        ...

    status = IcmpSendEcho2(pData->icmp_socket.sh, NULL,
                           g_pfnIcmpCallback, pong,
                           dst, reqdata, (WORD)reqsize, &opts,
                           pong->buf, (DWORD)pong->bufsize,
                           5 * 1000 /* ms */); // [2]
        ...
```

Heap-Overflow Exploitation

- We can control `pongsize` to overwrite next heap chunk with arbitrary size and arbitrary content
- We found that there is `struct socket` that we can overwrite, by overwriting this object we can control the program execution
- `struct socket` will be created everytime we create TCP/IP connection to the outside. We can just create a bunch of ICMP requests, to spray `struct socket` object in heap

```
struct socket
{
    struct socket   *so_next;
    struct socket   *so_prev;    /* For a linked list of sockets */

#if !defined(RT_OS_WINDOWS)
    int s;                       /* The actual socket */
#else
    union {
        int s;
        HANDLE sh;
    };
    uint64_t so_icmp_id; /* XXX: hack */
    uint64_t so_icmp_seq; /* XXX: hack */
#endif

    /* XXX union these with not-yet-used sbuf params */
    struct mbuf     *so_m; /* Pointer to the original SYN packet,
                            * for non-blocking connect()'s, and
                            * PING reply's */
    ...
};
```

Heap-Overflow Exploitation

```
#if defined(RT_OS_WINDOWS)
void slirp_select_poll(PNATState pData, int fTimeout)
#else /* RT_OS_WINDOWS */
void slirp_select_poll(PNATState pData, struct pollfd *polls, int
ndfs)
#endif /* !RT_OS_WINDOWS */
{
    struct socket *so, *so_next;
    int ret;
#if defined(RT_OS_WINDOWS)
    WSANETWORKEVENTS NetworkEvents;
    int rc;
    int error;
#endif
    ...
#if defined(RT_OS_WINDOWS)
        icmpwin_process(pData); // [1]
#else
    if (   (pData->icmp_socket.s != -1)
         && CHECK_FD_SET(&pData->icmp_socket, ignored, readfds))
        sorecvfrom(pData, &pData->icmp_socket);
#endif
    /*
     * Check TCP sockets
     */
    QSOCKET_FOREACH(so, so_next, tcp) // [2]
        /* { */
        Assert(!so->fUnderPolling);
    so->fUnderPolling = 1;
    if (slirpVerifyAndFreeSocket(pData, so)) // [3]
        CONTINUE(tcp);
}
```

- There is a function which collect `struct socket` object for handling events e.g: there is ICMP reply coming back
- In **[1]** it will process ICMP reply and heap overflow will happened on some `struct object`
- In **[2]** will collecting `struct socket` object including the one we overwritten
- An overwritten `struct socket` will be processed in **[3]**

Heap-Overflow Exploitation

- We control and set `fShouldBeRemoved` to 1 so it will call `sofree` with our controlled `pSocket` object.
- Then we call `m_freem` with pointer `pSocket->so_m` (mbuf object) controlled
- VRAM address is always spans in range `0xcb10000-0x1322000`, so it's safe to use `0x10000000` as our fake mbuf object for `so_m`
- Then it will call `m_freem -> m_free -> uma_zfree -> uma_zfree_arg -> slirp_uma_free`

```
static int slirpVerifyAndFreeSocket(PNATState pData, struct socket
*pSocket)
{
    AssertPtrReturn(pData, 0);
    AssertPtrReturn(pSocket, 0);
    AssertReturn(pSocket->fUnderPolling, 0);
    if (pSocket->fShouldBeRemoved) //[4]
    {
        pSocket->fUnderPolling = 0;
        sofree(pData, pSocket);
        /* pSocket is PHANTOM, now */
        return 1;
    }
    return 0;
}
```

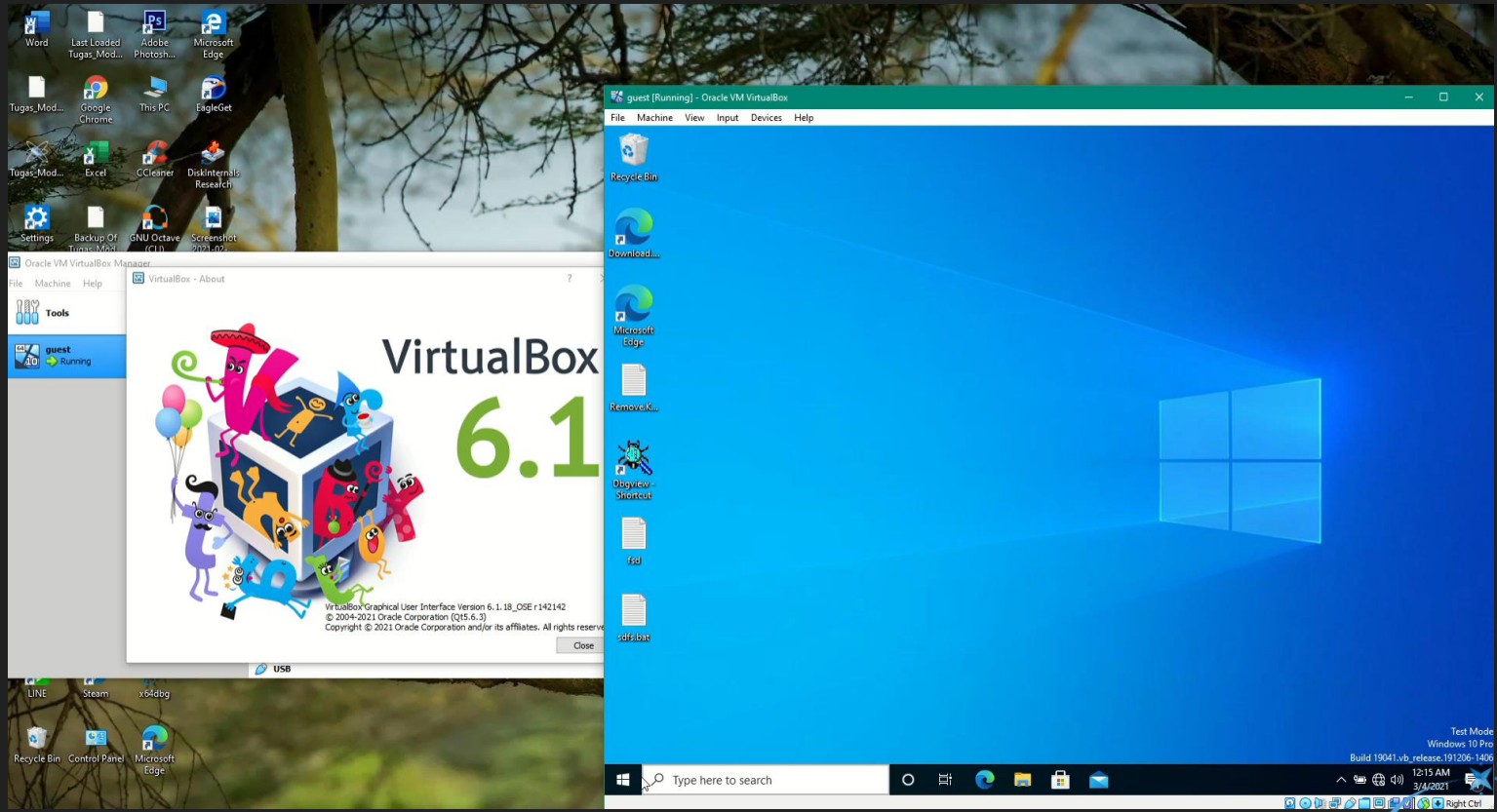
```
void
sofree(PNATState pData, struct socket *so)
{
    LogFlowFunc(("ENTER:%R[natsock]\n", so));
        ...
    /* check if mbuf haven't been already freed  */
    if (so->so_m != NULL)
    {
        m_freem(pData, so->so_m); //[5]
        so->so_m = NULL;
    }
}
```

Heap-Overflow Exploitation

```
static void slirp_uma_free(void *item, int size, uint8_t flags)
{
  struct item *it;
  uma_zone_t zone;
  ...
  it = &((struct item *)item)[-1];
  zone = it->zone;
  /* check border magic */
  ...
  if (zone->pfFini)
  {
  zone->pfFini(zone->pData, item, (int /*sigh*/)zone->size); //[6]
  }
}
```

- We can control `item`, and redirect program execution in **[6]**
- From the VBoxDD address we retrieve with CVE-2021-2321. We redirect address to stack pivot gadget, and execute our RopChain we already put at VRAM address
- Then the RopChain will prepare the shellcode and will execute `calc.exe`!

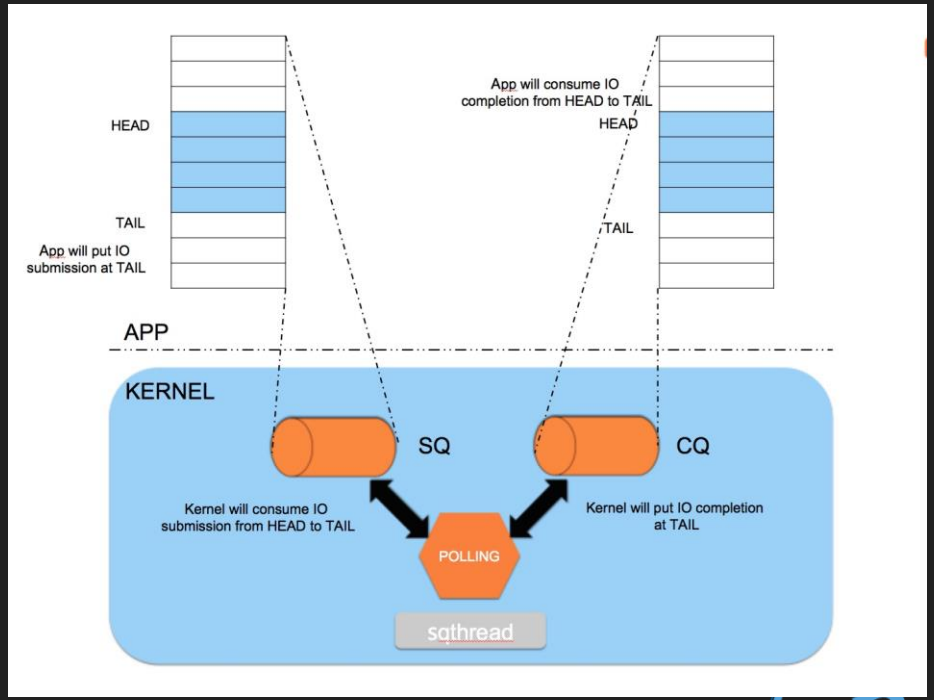
Demo


guest [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Oracle VM VirtualBox Manager
File Machine Help
Tools
guest
Running
VirtualBox - About
VirtualBox
6.1
VirtualBox Graphical User Interface Version 6.1.18_OSE r142142
© 2004-2021 Oracle Corporation (Qt5.6.3)
Copyright © 2021 Oracle Corporation and/or its affiliates. All rights reserved.
Close
USB
Recycle Bin
Type here to search
Test Mode
Windows 10 Pro
Build 19041.vb_release.191206-1406
12:15 AM
3/4/2021
Right Ctrl

Linux Kernel Privilege Escalation

Linux io_uring

- A new asynchronous I/O API start from kernel version 5.1
- Provide a pair of queue ring buffer shared between the kernel and userspace
- APP produce IO submission on SQ for kernel to consume.
- Kernel produce IO completion on SQ for APP to consume.



Linux io_uring

- It provides three system calls

```
/** * io_uring_setup - setup a context for performing asynchronous I/O */
int io_uring_setup(u32 entries, struct io_uring_params *p);

/** * io_uring_enter - initiate and/or complete asynchronous I/O */
int io_uring_enter(int fd, unsigned int to_submit, unsigned int min_complete,
    unsigned int flags, sigset_t *sig)

/** * io_uring_register - register files or user buffers for asynchronous I/O */
int io_uring_register(int fd, unsigned int opcode, void *arg,
    unsigned int nr_args)
```

Linux io_uring

It supports many asynchronous operations

- NOP only generates a completion event
- READV / WRITEV submit readv / writev operations
- PROVIDE_BUFFERS support for automatic buffers

```
enum {
    IORING_OP_NOP,
    IORING_OP_READV,
    IORING_OP_WRITEV,
    IORING_OP_FSYNC,
    IORING_OP_READ_FIXED,
    IORING_OP_WRITE_FIXED,
    ...
    IORING_OP_PROVIDE_BUFFERS,
    IORING_OP_REMOVE_BUFFERS,
    IORING_OP_TEE,

    /* this goes last, obviously */
    IORING_OP_LAST,
};
```

Vulnerability Details

When we do PROVIDE_BUFFERS operation, it will create a “large” automatic buffer (0x80000000,0x10)

```
static int io_provide_buffers_prep(struct io_kiocb *req,
                                   const struct io_uring_sqe *sqe)
{
    struct io_provide_buf *p = &req->pbuf;
    u64 tmp;

    if (sqe->ioprio || sqe->rw_flags)
        return -EINVAL;

    tmp = READ_ONCE(sqe->fd);
    if (!tmp || tmp > USHRT_MAX)
        return -E2BIG;
    p->nbufs = tmp; // __u16
    p->addr = READ_ONCE(sqe->addr);
    p->len = READ_ONCE(sqe->len); //__s32

    if (!access_ok(u64_to_user_ptr(p->addr), (p->len * p->nbufs))) //multiplication overflow
        return -EFAULT;
    ...
```

Vulnerability Details

When we do READV operation with our “large” automatic buffers, our final read length will larger than MAX_RW_COUNT(0x7ffff000)

```
static struct io_buffer *io_buffer_select(struct io_kiocb *req, size_t *len,
                                          int bgid, struct io_buffer *kbuf,
                                          bool needs_lock)
{
    ...
        if (*len > kbuf->len)
            *len = kbuf->len;
    } else {
        kbuf = ERR_PTR(-ENOBUFS);
    }

    io_ring_submit_unlock(req->ctx, needs_lock);

    return kbuf;
```

Vulnerability Details

When we do read on /proc/self/mem, it could cause heap overflow

```
static ssize_t mem_rw(struct file *file, char __user *buf,
                      size_t count, loff_t *ppos, int write)
{

...
    page = (char *)__get_free_page(GFP_KERNEL);
...

    while (count > 0) {
        int this_len = min_t(int, count, PAGE_SIZE); // int compare with size_t
...
        this_len = access_remote_vm(mm, addr, page, this_len, flags); // this cause overflow on "page"
        if (!this_len) {
            if (!copied)
                copied = -EIO;
            break;
        }
```

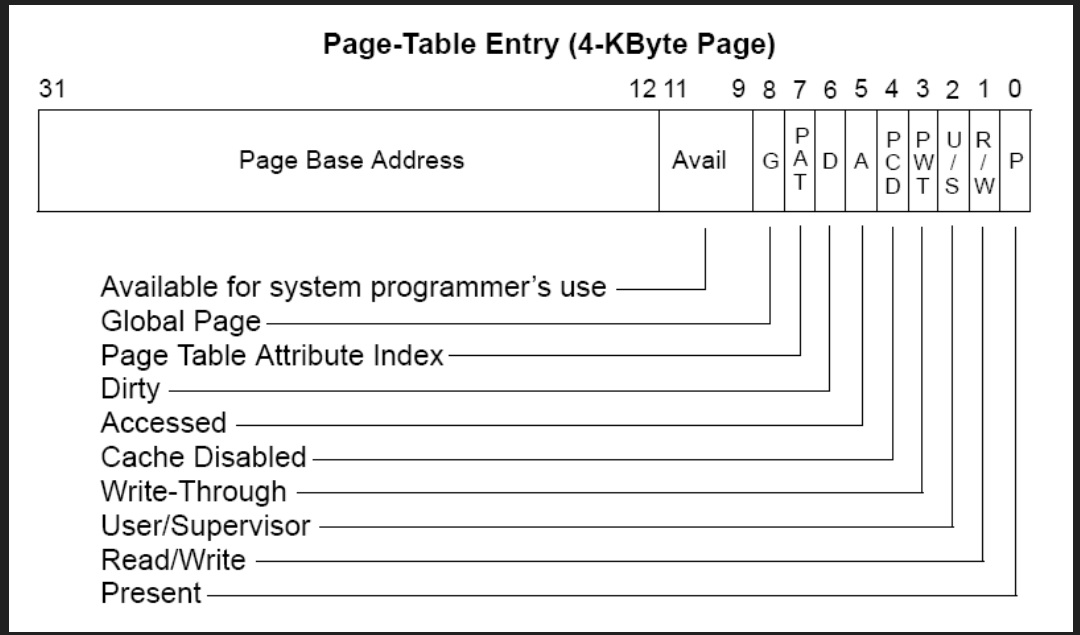
The concept of exploitation

- Step
  - IORING_OP_PROVIDE_BUFFERS to provide a buffer with length 0x80000000
  - Open “/proc/self/mem” and lseek to a readable address
  - IORING_OP_READV to cause heap overflow

- Ability
  - We can choose where to start read (control content)
  - access_remote_vm will stop copy at invalid address (control overflow length)
  - We can control the length and value we want to overflow

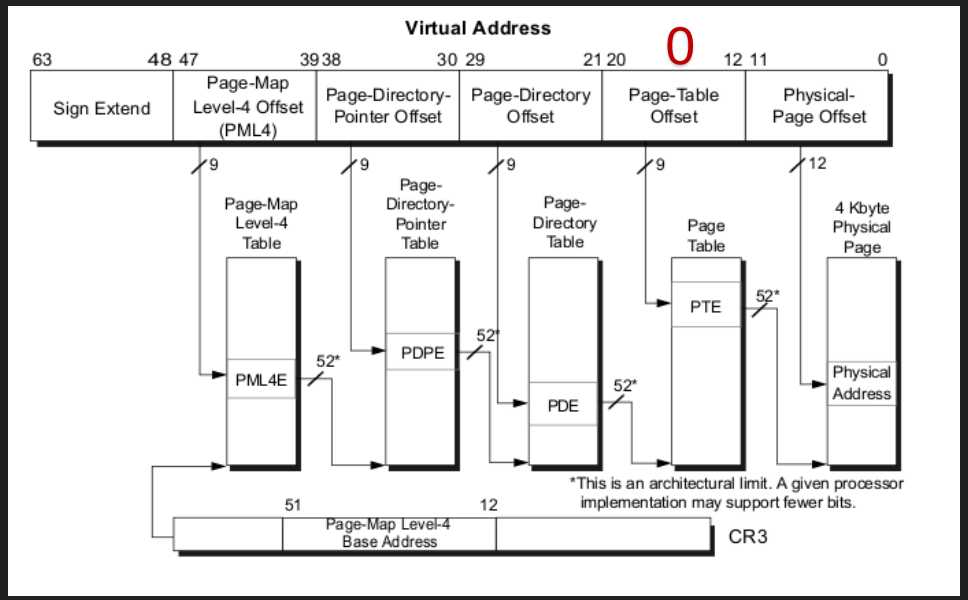
Overwrite the page table entry

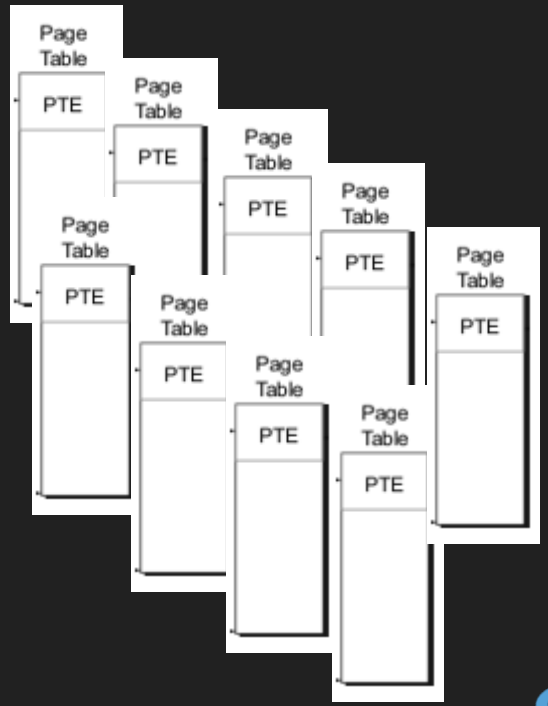
Overwrite write bit to make read-only page writeable



Effective spray

Create a lot of page table with only one entry





Example Code

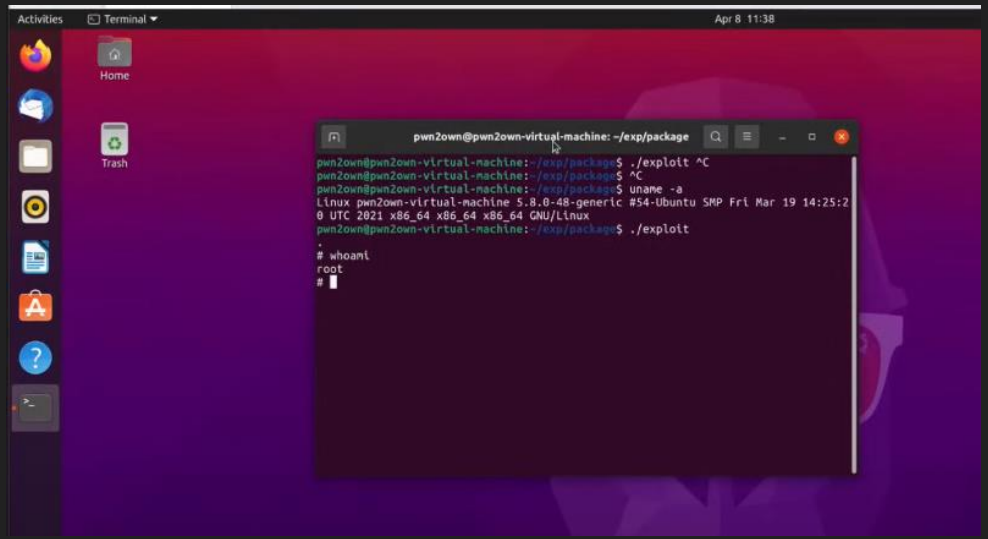
```
int fd = open("/bin/passwd", 0);

for (size_t i = 0; i < (1 << 9); i++)
    addr[i] = (void *)mmap((void *)( i << (9 + 12) ,
        0x1000, PROT_READ, MAP_SHARED | MAP_FIXED, fd, 0x0);

char x = 0;
for (int i = 0; i < (1 << 9); i++)
    x += addr[i][0]; //actually create page table for access
```

Local Privilege Escalation

- Spray a lot of read-only pages which map to a suid program (ex. /bin/passwd)
- Using io_uring to overflow a byte to enable writable bit
- Using system call to test those readonly page is writable
- Overwrite /bin/passwd with our shellcode
- Get Root



Thank You