## A deep-dive into the SolarWinds Serv-U SSH vulnerability

microsoft.com/security/blog/2021/09/02/a-deep-dive-into-the-solarwinds-serv-u-ssh-vulnerability/

September 2, 2021

Several weeks ago, Microsoft detected a 0-day remote code execution exploit being used to attack the SolarWinds Serv-U FTP software in limited and targeted attacks. The Microsoft Threat Intelligence Center (MSTIC) attributed the <u>attack</u> with high confidence to DEV-0322, a group operating out of China, based on observed victimology, tactics, and procedures. In this blog, we share technical information about the vulnerability, tracked as <u>CVE-2021-35211</u>, that we shared with SolarWinds, who promptly released <u>security updates</u> to fix the vulnerability and mitigate the attacks.

This analysis was conducted by the Microsoft Offensive Research & Security Engineering team, a focused group tasked with supporting teams like MSTIC with exploit development expertise. Our team's remit is to make computing safer. We do this by leveraging our knowledge of attacker techniques and processes to build and improve protections in Windows and Azure through reverse engineering, attack creation and replication, vulnerability research, and intelligence sharing.

In early July, MSTIC provided our team with data that seemed to indicate exploit behavior against a newly-discovered vulnerability in the SolarWinds Serv-U FTP server's SSH component. Although the intel contained useful indicators, it lacked the exploit in question, so our team set out to reconstruct the exploit, which required to first find and understand the new vulnerability in the Serv-U SSH-related code.

As we knew this was a remote, pre-auth vulnerability, we quickly constructed a fuzzer focused on the pre-auth portions of the SSH handshake and noticed that the service captured and passed all access violations without terminating the process. It immediately became evident that the Serv-U process would make stealthy, reliable exploitation attempts simple to accomplish. We concluded that the exploited vulnerability was caused by the way Serv-U initially created an OpenSSL AES128-CTR context. This, in turn, could allow the use of uninitialized data as a function pointer during the decryption of successive SSH messages. Therefore, an attacker could exploit this vulnerability by connecting to the open SSH port and sending a malformed pre-auth connection request. We also discovered that the attackers were likely using DLLs compiled without address space layout randomization (ASLR) loaded by the Serv-U process to facilitate exploitation.

We shared these findings, as well as the fuzzer we created, with SolarWinds through <u>Coordinated Vulnerability Disclosure</u> (CVD) via <u>Microsoft Security Vulnerability Research</u> (MSVR), and worked with them to fix the issue. This is an example of intelligence sharing

and industry collaboration that result in comprehensive protection for the broader community through detection of attacks through products and fixing vulnerabilities through security updates.

#### Vulnerability in Serv-U's implementation of SSH

Secure Shell (SSH) is a widely adopted protocol for secure communications over an untrusted network. The protocol behavior is defined in multiple requests for comment (RFCs), and existing implementations are available in open-source code; we primarily used <u>RFC 4253</u>, <u>RFC 4252</u>, and <u>libssh</u> as references for this analysis.

The implementation of SSH in Serv-U was found by enumerating references to the "SSH-" string, which must be present in the first data sent to the server. The most likely instance of such code was the following:

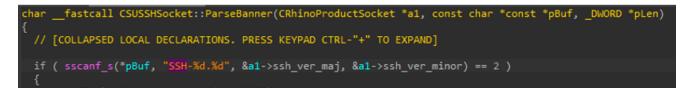


Figure 1. Promising instance of "SSH-" string

Putting a breakpoint on the above code and attempting to connect to Serv-U with an SSH client confirmed our hypothesis and resulted in the breakpoint being hit with the following call stack:

<pre># Child-SP 00 0000068`b68ff150 01 00000068`b68ff1e0 02 00000068`b68ff2b0 03 0000068`b68ff2f0 04 0000068`b68ff2f0</pre>	00000184`d3e58dd7 00000184`d3e58c5c 00000184`d3e56c4e	Call Site Serv_U!CUPnPNotifyEvent::SetTimeout+0x446e Serv_U!CUPnPNotifyEvent::SetTimeout+0x71dd RhinoNET!CRhinoSocket::ProcessReceiveBuffer+0x33 RhinoNET!CRhinoSocket::OnReceive+0x170
04 0000068`b68ff340 05 0000068`b68ff370 06 0000068`b68ff3a0 07 0000068`b68ff3d0	00000184`d3e5356b 00000184`d3e5350f	RhinoNET!CRhinoProductSocket::OnReceive+0x3e RhinoNET!CAsyncSocketX::DoCallBack+0x107 RhinoNET!CAsyncSocketX::ProcessAuxQueue+0x53 RhinoNET!CSocketWndX::OnSocketNotify+0x13

Figure 2. The call stack resulting from a break point set on code in Figure 1.

At this point, we noticed that *Serv-U.dll* and *RhinoNET.dll* both have ASLR support disabled, making them prime locations for ROP gadgets, as any addresses within them will be constant across any server instances running on the internet for a given Serv-U version.

After reversing related code in the *RhinoNET* and *Serv-U* DLLs, we could track SSH messages' paths as Serv-U processes them. To handle an incoming SSH connection, *Serv-U.dll* creates a *CSUSSHSocket* object, which is derived from the *RhinoNET!CRhinoSocket class*. The *CSUSSHSocket* object lifetime is the length of the TCP connection—it persists across possibly many individual TCP packets. The underlying *CRhinoSocket* provides a buffered interface to the socket such that a single TCP packet may contain any number of

bytes. This implies a single packet may include any number of SSH messages (provided they fit in the maximum buffer size), as well as partial SSH messages. The *CSUSSHSocket::ProcessRecvBuffer* function is then responsible for parsing the SSH messages from the buffered socket data.

*CSUSSHSocket::ProcessRecvBuffer* begins by checking for the SSH version with *ParseBanner*. If *ParseBanner* successfully parses the SSH version from the banner, *ProcessRecvBuffer* then loops over *ParseMessage*, which obtains a pointer to the current message in the socket data and extracts the *msg\_id* and *length* fields from the message (more on the *ParseMessage* function later).

```
if ( !a1->ssh banner_status )
 bUsed = CSUSSHSocket::ParseBanner(a1, (const char *const *)pBuf, p_buf_len);
if ( a1->ssh_banner_status == 1 )
 msg = CSocketMessage::create(a1, 5304);
 if (msg)
   v6 = 0;
   packet_len = 0;
  payload_len = 0;
   len_comp = 0;
  v27 = 0;
  v7 = 0;
  v25 = 0;
  pad_len = 0;
  msg id = 0;
   pkt = 0i64;
   v8 = 0;
   v22 = 0;
   buf len before = *p buf len;
   v24 = *p_buf_len;
   bUsed = 1;
   while ( bUsed
        && CSUSSHSocket::ParseMessage(
              a1,
             &bUsed,
              pBuf,
             p buf len,
              &packet_len,
             &payload len,
              &len comp,
              &pad len,
              &msg id,
              &pkt) )
```

Figure 3. Selection of code from CSUSSHSocket::ProcessRecvBuffer processing loop

The socket data being iterated over is conceptually an array of the pseudo-C structure *ssh\_msg\_t*, as seen below. The message data is contained within the payload buffer, the first byte of which is considered the *msg\_id*:

```
struct ssh_msg_t {
    u32 packet_length;
    u8 padding_length;
    // first byte is the msg_id
    u8 payload[packet_length - padding_length - 1];
    u8 random_padding[padding_length];
    // mac_length is 0 initially, until non-"none"
    // mac algorithm is negotiated
    u8 mac[mac_length];
};
```

*ProcessRecvBuffer* then dispatches handling of the message based on the *msg\_id*. Some messages are handled directly from the message parsing loop, while others get passed to *ssh\_pkt\_others*, which posts the message to a queue for another thread to pick up and process.

```
case 5u:
ssh_pkt_5_svc_req(a1, pkt, packet_len);
 break;
 if ( !ssh_pkt_20_kexinit(a1, (char *)pkt - 1, packet_len + 1) )
   goto LABEL_27;
 break:
ssh_pkt_21_newkeys(a1);
 break;
ssh pkt 30 kexdh init(a1, pkt, packet len);
 break:
 ssh_pkt_32(a1, pkt, packet_len);
 break:
case 34u:
 ssh_pkt_34(a1, pkt, packet_len, 0);
 break:
default:
  if ( a1->m_bCipherActive || msg_id && msg_id <= 49u )
    ssh_pkt_others(a1, &msg, pkt, payload_len, pad_len, msg_id, buf_len_read, v13);
    v21 = 1;
```

Figure 4.Pre-auth reachable handlers in CSUSSHSocket::ProcessRecvBuffer

If the *msg\_id* is deferred to the alternate thread, *CSSHSession::OnSSHMessage* processes it. This function mainly deals with messages that need to interact with Serv-U managed user profile data (e.g., authentication against per-user credentials) and UI updates. *CSSHSession::OnSSHMessage* turned out to be uninteresting in terms of vulnerability hunting as most message handlers within it require successful user authentication (initial telemetry indicated this was a pre-authentication vulnerability), and no vulnerabilities were found in the remaining handlers.

When initially running fuzzers against Serv-U with a debugger attached, it was evident that the application was catching exceptions which would normally crash a process (such as access violations), logging the error, modifying state just enough to avoid termination of the process, and then continuing as if there had been no problem. This behavior improves uptime of the file server application but also results in possible memory corruption lingering around in the process and building up over time. As an attacker, this grants opportunities like brute-forcing addresses of code or data with dynamic addresses.

This squashing of access violations assists with exploitation, but for fuzzing, we filtered out "uninteresting" exceptions generated by read/write access violations and let the fuzzer run until hitting a fault wherein RIP had been corrupted. This quickly resulted in the following crashing context:

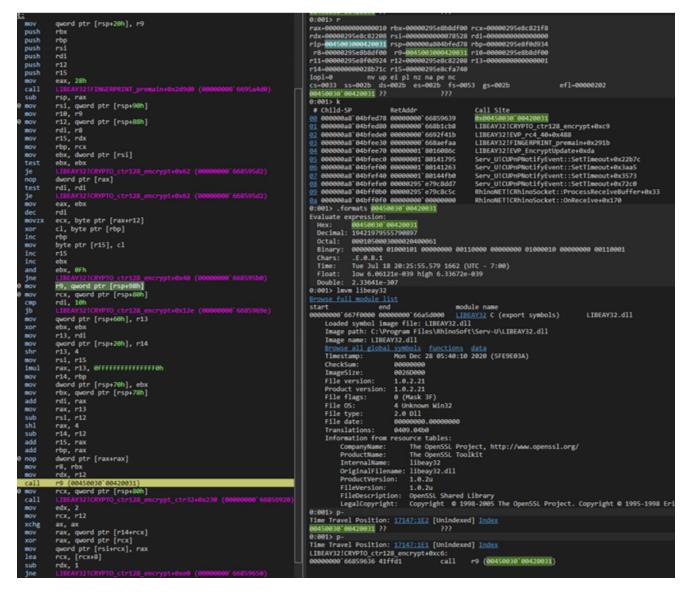


Figure 5. WinDbg showing crashing context from fuzzer-generated SSH messages

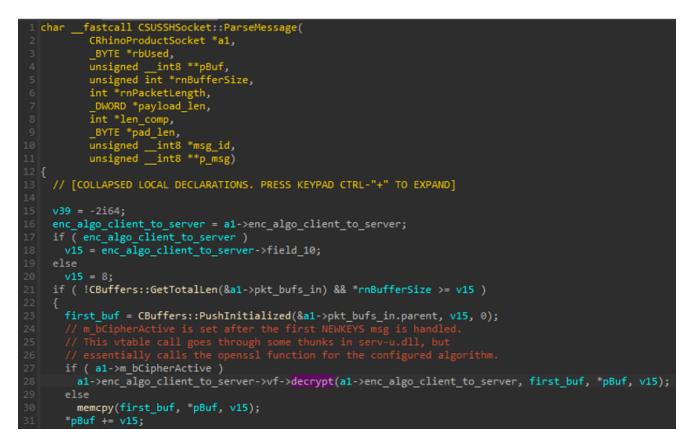
As seen above, *CRYPTO\_ctr128\_encrypt* in *libeay32.dll* (part of OpenSSL) attempted to call an invalid address. The version of OpenSSL used is 1.0.2u, so we obtained the <u>sources</u> to peruse. The following shows the relevant OpenSSL function:

```
void CRYPTO ctr128 encrypt(const unsigned char *in, unsigned char *out,
                           size t len, const void *key,
                           unsigned char ivec[16],
                           unsigned char ecount buf[16], unsigned int *num,
                           block128 f block)
   snip */
                (*block) (ivec, ecount buf, key);
  snip */
static int aes_ctr_cipher(EVP_CIPHER_CTX *ctx, unsigned char *out,
                          const unsigned char *in, size t len)
{
    unsigned int num = ctx->num;
    EVP AES KEY *dat = (EVP AES KEY *) ctx->cipher data;
    if (dat->stream.ctr)
        CRYPTO_ctr128_encrypt_ctr32(in, out, len, &dat->ks,
                                    ctx->iv, ctx->buf, &num, dat->stream.ctr);
    else
        CRYPTO ctr128 encrypt(in, out, len, &dat->ks,
                              ctx->iv, ctx->buf, &num, dat->block);
    ctx->num = (size t)num;
    return 1;
```

Meanwhile, the following shows the structure that is passed:

```
typedef void (*block128_f)(const unsigned char in[16],
    unsigned char out[16],
    const void *key);
typedef struct {
    union {
        double align;
        AES_KEY ks;
    } ks;
    block128_f block;
    union {
        cbc128_f cbc;
        ctr128_f ctr;
    } stream;
} EVP AES KEY;
```

The crashing function was reached from the OpenSSL API boundary via the following path: EVP\_EncryptUpdate -> evp\_EncryptDecryptUpdate -> aes\_ctr\_cipher -> CRYPTO\_ctr128\_encrypt. Looking further up the call stack, it is evident that Serv-U calls *EVP\_EncryptUpdate* from *CSUSSHSocket::ParseMessage*, as seen below:



*Figure 6. Location of call into OpenSSL, wherein attacker-controlled function pointer may be invoked* 

At this point, we manually minimized the TCP packet buffer produced by the fuzzer until only the SSH messages required to trigger the crash remained. In notation like that used in the RFCs, the required SSH messages were:

```
Version message
SSH MSG KEXINIT
    "ecdh-sha2-nistp256"
    "ssh-rsa"
    "aes128-ctr"
    "none"
    "none"
    "none"
    "none"
    "none"
    .....
    .....
    false
    0
SSH MSG NEWKEYS
SSH MSG USERAUTH_REQUEST
    "none"
```

Note that the following description references "encrypt" functions being called when the crashing code path is clearly attempting to decrypt a buffer. This is not an error: Serv-U uses the encrypt OpenSSL API and, while not optimal for code clarity, it is behaviorally correct since Advanced Encryption Standard (AES) is operating in counter (CTR) mode.

After taking a Time Travel Debugging trace and debugging through the message processing sequence, we found that the root cause of the issue was that Serv-U initially creates the OpenSSL AES128-CTR context with code like the following:

```
EVP_CIPHER_CTX* ctx = EVP_CIPHER_CTX_new();
// |key| and |iv| set to NULL
EVP_EncryptInit_ex(ctx, EVP_aes_128_ctr(), NULL, NULL, NULL);
```

Calling <u>EVP\_EncryptInit\_ex</u> with NULL key and/or IV is valid, and Serv-U does so in this case because the context is created while handling the KEXINIT message, which is before key material is ready. However, AES key expansion is not performed until the key is set, and the data in the *ctx->cipher\_data* structure remains uninitialized until the key expansion is performed. We can (correctly) surmise that our sequence of messages to hit the crash has caused *enc\_algo\_client\_to\_server->decrypt* to be called before the key material is initialized. The Serv-U KEXINIT handler creates objects for all parameters given in the message. However, the corresponding objects currently active for the connection are not replaced with the newly created ones until the following NEWKEYS message is processed. The client always completes the key exchange process In a normal SSH connection before issuing a NEWKEYS message. Serv-U processed NEWKEYS (thus setting the *m\_bCipherActive* flag and replacing the cipher objects) no matter the connection state or key exchange. From this, we can see that the last message type in our fuzzed sequence does not matter—there only needs to be some data remaining to be processed in the socket buffer to trigger decryption after the partially initialized AES CTR cipher object has been activated.

### Exploitation

As the vulnerability allows loading RIP from uninitialized memory and as there are some modules without ASLR in the process, exploitation is not so complicated: we can find a way to control the content of the uninitialized *cipher\_data* structure, point the *cipher\_data->block* function pointer at some initial ROP gadget, and start a ROP chain. Because of the exception handler causing any fault to be ignored, we do not necessarily need to attain reliable code execution upon the first packet. It is possible to retry exploitation until code execution is successful, however this will leave traces in log files and as such it may be worthwhile to invest more effort into a different technique which would avoid logging. The first step is to find the size of the *cipher\_data* allocation, as the most direct avenue to prefill the buffer is to spray allocations of the target allocation size and free them before attempting to reclaim the address as *cipher\_data*. *ctx->cipher\_data* is allocated and assigned in EVP\_CipherInit\_ex with the following line:

ctx->cipher\_data = OPENSSL\_malloc(ctx->cipher->ctx\_size);

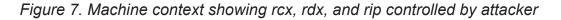
With a debugger, we can see the *ctx\_size* in our case is *0x108*, and that this allocator winds up calling *ucrtbase!\_malloc\_base*. From previous reversing, we know that both *CRhinoSocket* and *CSUSSHSocket* levels of packet parsing call *operator new[]* to allocate space to hold the packets we send. Luckily, that also winds up in *ucrtbase!\_malloc\_base*, using the same heap. Therefore, prefilling the target allocation is as simple as sending a properly sized TCP packet or SSH message and then closing the connection to ensure it is freed. Using this path to spray does not trigger other allocations of the same size, so we don't have to worry about polluting the heap.

Another important value to pull out of the debugger/disassembly is *offsetof(EVP\_AES\_KEY, block)*, as that offset in the sprayed data needs to be set to the initial ROP gadget. This value is *0xf8*. Conveniently, most of the rest of the *EVP\_AES\_KEY* structure can be used for the ROP chain contents itself, and a pointer to the base of this structure exists in registers *rbx*, *r8*, and *r10* at the time of the controlled function pointer call.

As a simple proof of concept, consider the following python code:

```
alloc_size = 0x108
rop = bytearray(alloc_size)
def write_qw(0, qw): rop[0:0+8] = struct.pack('<Q', qw)
write_qw(0x10, 0x41 * 0x010101010101010)
write_qw(0x38, 0x42 * 0x0101010101010101)
write_qw(0x58, 0x43 * 0x0101010101010101)
write_qw(0x58, 0x1800E19EC)
spray = struct.pack('>IBBH', alloc_size, 5, 99, 0) + rop
for _ in range(20):
    send_packet(b'SSH-2.0-\r\n' + spray)
send packet(fuzzer trigger)
```

The above results in the following context in the debugger:



# Conclusion: Responsible disclosure and industry collaboration improves security for all

Our research shows that the Serv-U SSH server is subject to a pre-auth remote code execution vulnerability that can be easily and reliably exploited in the default configuration. An attacker can exploit this vulnerability by connecting to the open SSH port and sending a

malformed pre-auth connection request. When successfully exploited, the vulnerability could then allow the attacker to install or run programs, such as in the case of the targeted attack we previously reported.

We shared our findings to SolarWinds through <u>Coordinated Vulnerability Disclosure</u> (CVD). We also shared the fuzzer we created. SolarWinds released an <u>advisory and security patch</u>, which we strongly encourage customers to apply. If you are not sure if your system is affected, open a support case in the <u>SolarWinds Customer Portal</u>.

In addition to sharing vulnerability details and fuzzing tooling with SolarWinds, we also recommended <u>enabling ASLR compatibility</u> for all binaries loaded in the Serv-U process. Enabling ASLR is a simple compile-time flag which is enabled by default and has been available since Windows Vista. ASLR is a critical security mitigation for services which are exposed to untrusted remote inputs, and requires that all binaries in the process are compatible in order to be effective at preventing attackers from using hardcoded addresses in their exploits, as was possible in Serv-U.

We would like to thank SolarWinds for their prompt response. This case further underscores the need for constant collaboration among software vendors, security researchers, and other players to ensure the safety and security of users' computing experience.

#### Microsoft Offensive Research & Security Engineering team