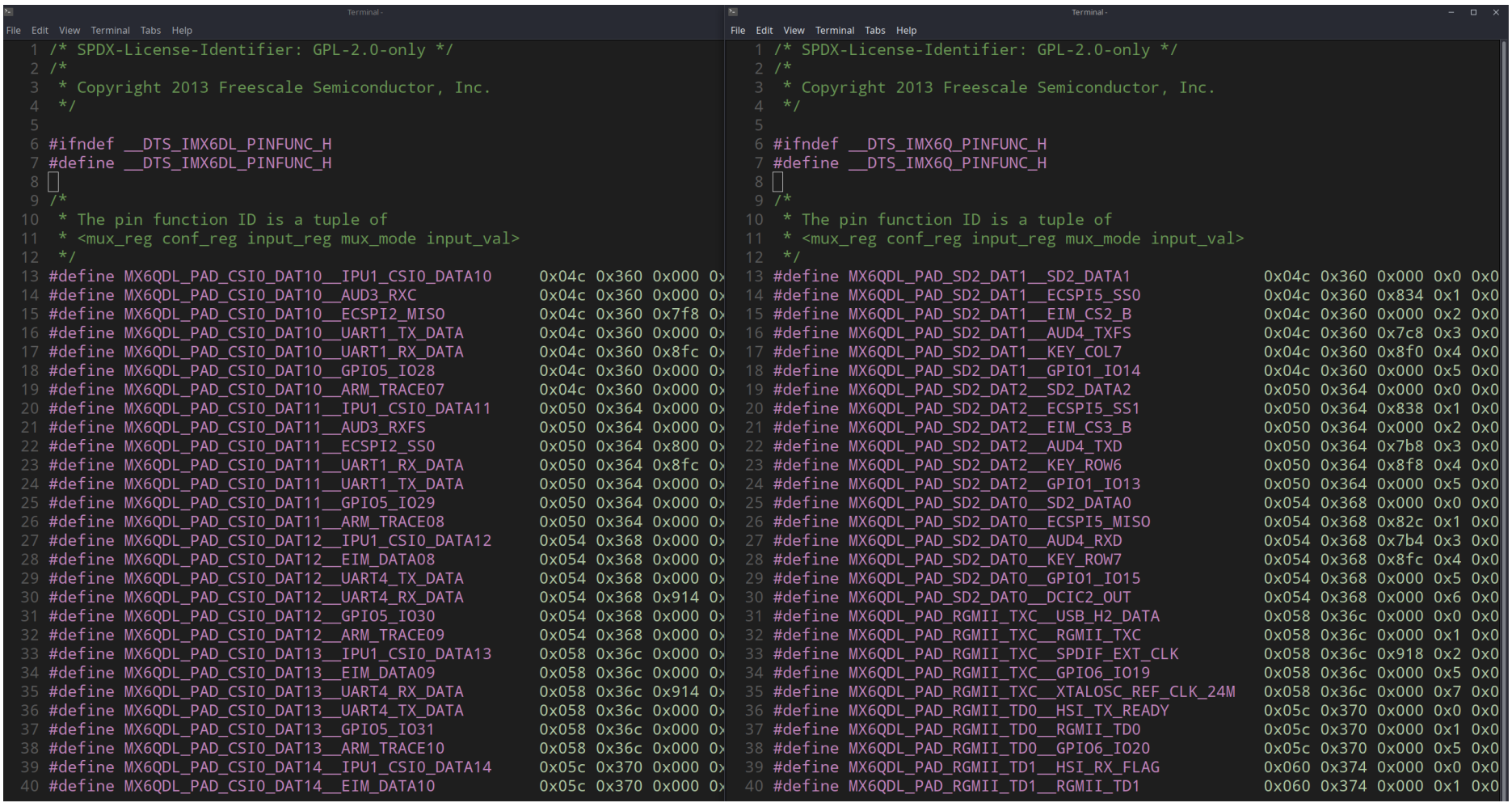
从kernel.org下载最新的linux 5.18.1源代码，项目内File粒度克隆检测结果（开头一小部分）：

```
1,384,1,383
1,428,1,425
1,524,1,272
1,572,1,569
1,864,1,859
1,1129,1,1113
1,1409,1,1408
1,1669,1,1661
1,1677,1,1676
1,1793,1,1458
1,1927,1,1926
1,1927,1,1924
1,2005,1,1851
1,2046,1,2043
1,2055,1,413
1,2061,1,423
1,2062,1,424
1,2145,1,524
1,2198,1,583
1,2244,1,632
1,2481,1,357
1,2550,1,632
```

最初的检测结果是大量头文件的相似，例如`arch/arm/boot/dts/imx6dl-pinfunc.h`与`arm/boot/dts/imx6q-pinfunc.h`：



```
/* SPDX-License-Identifier: GPL-2.0-only */
/*
 * Copyright 2013 Freescale Semiconductor, Inc.
 */

#ifndef __DTS_IMX6DL_PINFUNC_H
#define __DTS_IMX6DL_PINFUNC_H

/*
 * The pin function ID is a tuple of
 * <mux_reg conf_reg input_reg mux_mode input_val>
 */
#define MX6QDL_PAD_CSI0_DAT10__IPU1_CSI0_DATA10      0x04c 0x360 0x000 0x
#define MX6QDL_PAD_CSI0_DAT10__AUD3_RXC              0x04c 0x360 0x000 0x
#define MX6QDL_PAD_CSI0_DAT10__ECSPI2_MISO           0x04c 0x360 0x7f8 0x
#define MX6QDL_PAD_CSI0_DAT10__UART1_TX_DATA         0x04c 0x360 0x000 0x
#define MX6QDL_PAD_CSI0_DAT10__UART1_RX_DATA         0x04c 0x360 0x8fc 0x
#define MX6QDL_PAD_CSI0_DAT10__GPIO5_IO28            0x04c 0x360 0x000 0x
#define MX6QDL_PAD_CSI0_DAT10__ARM_TRACE07           0x04c 0x360 0x000 0x
#define MX6QDL_PAD_CSI0_DAT11__IPU1_CSI0_DATA11      0x050 0x364 0x000 0x
#define MX6QDL_PAD_CSI0_DAT11__AUD3_RXFS             0x050 0x364 0x000 0x
#define MX6QDL_PAD_CSI0_DAT11__ECSPI2_SS0            0x050 0x364 0x800 0x
#define MX6QDL_PAD_CSI0_DAT11__UART1_RX_DATA         0x050 0x364 0x8fc 0x
#define MX6QDL_PAD_CSI0_DAT11__UART1_TX_DATA         0x050 0x364 0x000 0x
#define MX6QDL_PAD_CSI0_DAT11__GPIO5_IO29            0x050 0x364 0x000 0x
#define MX6QDL_PAD_CSI0_DAT11__ARM_TRACE08           0x050 0x364 0x000 0x
#define MX6QDL_PAD_CSI0_DAT12__IPU1_CSI0_DATA12      0x054 0x368 0x000 0x
#define MX6QDL_PAD_CSI0_DAT12__EIM_DATA08            0x054 0x368 0x000 0x
#define MX6QDL_PAD_CSI0_DAT12__UART4_TX_DATA         0x054 0x368 0x000 0x
#define MX6QDL_PAD_CSI0_DAT12__UART4_RX_DATA         0x054 0x368 0x914 0x
#define MX6QDL_PAD_CSI0_DAT12__GPIO5_IO30            0x054 0x368 0x000 0x
#define MX6QDL_PAD_CSI0_DAT12__ARM_TRACE09           0x054 0x368 0x000 0x
#define MX6QDL_PAD_CSI0_DAT13__IPU1_CSI0_DATA13      0x058 0x36c 0x000 0x
#define MX6QDL_PAD_CSI0_DAT13__EIM_DATA09            0x058 0x36c 0x000 0x
#define MX6QDL_PAD_CSI0_DAT13__UART4_RX_DATA         0x058 0x36c 0x914 0x
#define MX6QDL_PAD_CSI0_DAT13__UART4_TX_DATA         0x058 0x36c 0x000 0x
#define MX6QDL_PAD_CSI0_DAT13__GPIO5_IO31            0x058 0x36c 0x000 0x
#define MX6QDL_PAD_CSI0_DAT13__ARM_TRACE10           0x058 0x36c 0x000 0x
#define MX6QDL_PAD_CSI0_DAT14__IPU1_CSI0_DATA14      0x05c 0x370 0x000 0x
#define MX6QDL_PAD_CSI0_DAT14__EIM_DATA10            0x05c 0x370 0x000 0x
```

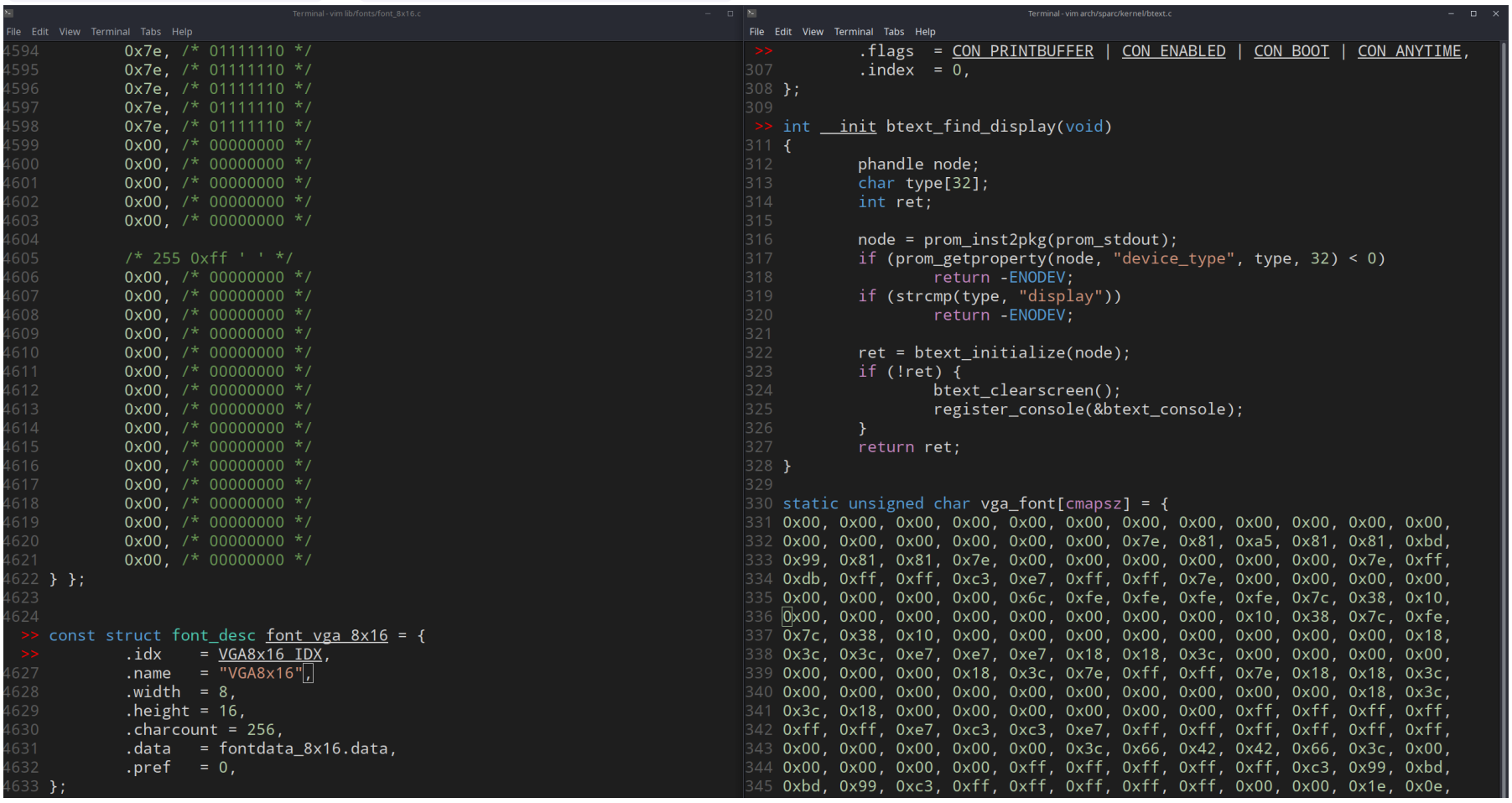
```
/* SPDX-License-Identifier: GPL-2.0-only */
/*
 * Copyright 2013 Freescale Semiconductor, Inc.
 */

#ifndef __DTS_IMX6Q_PINFUNC_H
#define __DTS_IMX6Q_PINFUNC_H

/*
 * The pin function ID is a tuple of
 * <mux_reg conf_reg input_reg mux_mode input_val>
 */
#define MX6QDL_PAD_SD2_DAT1__SD2_DATA1               0x04c 0x360 0x000 0x0 0x0
#define MX6QDL_PAD_SD2_DAT1__ECSPI5_SS0              0x04c 0x360 0x834 0x1 0x0
#define MX6QDL_PAD_SD2_DAT1__EIM_CS2_B               0x04c 0x360 0x000 0x2 0x0
#define MX6QDL_PAD_SD2_DAT1__AUD4_TXFS               0x04c 0x360 0x7c8 0x3 0x0
#define MX6QDL_PAD_SD2_DAT1__KEY_COL7                0x04c 0x360 0x8f0 0x4 0x0
#define MX6QDL_PAD_SD2_DAT1__GPIO1_IO14              0x04c 0x360 0x000 0x5 0x0
#define MX6QDL_PAD_SD2_DAT2__SD2_DATA2               0x050 0x364 0x000 0x0 0x0
#define MX6QDL_PAD_SD2_DAT2__ECSPI5_SS1              0x050 0x364 0x838 0x1 0x0
#define MX6QDL_PAD_SD2_DAT2__EIM_CS3_B               0x050 0x364 0x000 0x2 0x0
#define MX6QDL_PAD_SD2_DAT2__AUD4_TXD                0x050 0x364 0x7b8 0x3 0x0
#define MX6QDL_PAD_SD2_DAT2__KEY_ROW6                0x050 0x364 0x8f8 0x4 0x0
#define MX6QDL_PAD_SD2_DAT2__GPIO1_IO13              0x050 0x364 0x000 0x5 0x0
#define MX6QDL_PAD_SD2_DAT0__SD2_DATA0               0x054 0x368 0x000 0x0 0x0
#define MX6QDL_PAD_SD2_DAT0__ECSPI5_MISO             0x054 0x368 0x82c 0x1 0x0
#define MX6QDL_PAD_SD2_DAT0__AUD4_RXD                0x054 0x368 0x7b4 0x3 0x0
#define MX6QDL_PAD_SD2_DAT0__KEY_ROW7                0x054 0x368 0x8fc 0x4 0x0
#define MX6QDL_PAD_SD2_DAT0__GPIO1_IO15              0x054 0x368 0x000 0x5 0x0
#define MX6QDL_PAD_SD2_DAT0__DCIC2_OUT               0x054 0x368 0x000 0x6 0x0
#define MX6QDL_PAD_RGMII_TXC__USB_H2_DATA            0x058 0x36c 0x000 0x0 0x0
#define MX6QDL_PAD_RGMII_TXC__RGMII_TXC              0x058 0x36c 0x000 0x1 0x0
#define MX6QDL_PAD_RGMII_TXC__SPDIF_EXT_CLK          0x058 0x36c 0x918 0x2 0x0
#define MX6QDL_PAD_RGMII_TXC__GPIO6_IO19             0x058 0x36c 0x000 0x5 0x0
#define MX6QDL_PAD_RGMII_TXC__XTALOSC_REF_CLK_24M    0x058 0x36c 0x000 0x7 0x0
#define MX6QDL_PAD_RGMII_TD0__HSI_TX_READY           0x05c 0x370 0x000 0x0 0x0
#define MX6QDL_PAD_RGMII_TD0__RGMII_TD0              0x05c 0x370 0x000 0x1 0x0
#define MX6QDL_PAD_RGMII_TD0__GPIO6_IO20             0x05c 0x370 0x000 0x5 0x0
#define MX6QDL_PAD_RGMII_TD1__HSI_RX_FLAG            0x060 0x374 0x000 0x0 0x0
#define MX6QDL_PAD_RGMII_TD1__RGMII_TD1              0x060 0x374 0x000 0x1 0x0
```

排除头文件再检测（文件id接近，他们的位置也接近，他们更可能是“合理克隆”）：

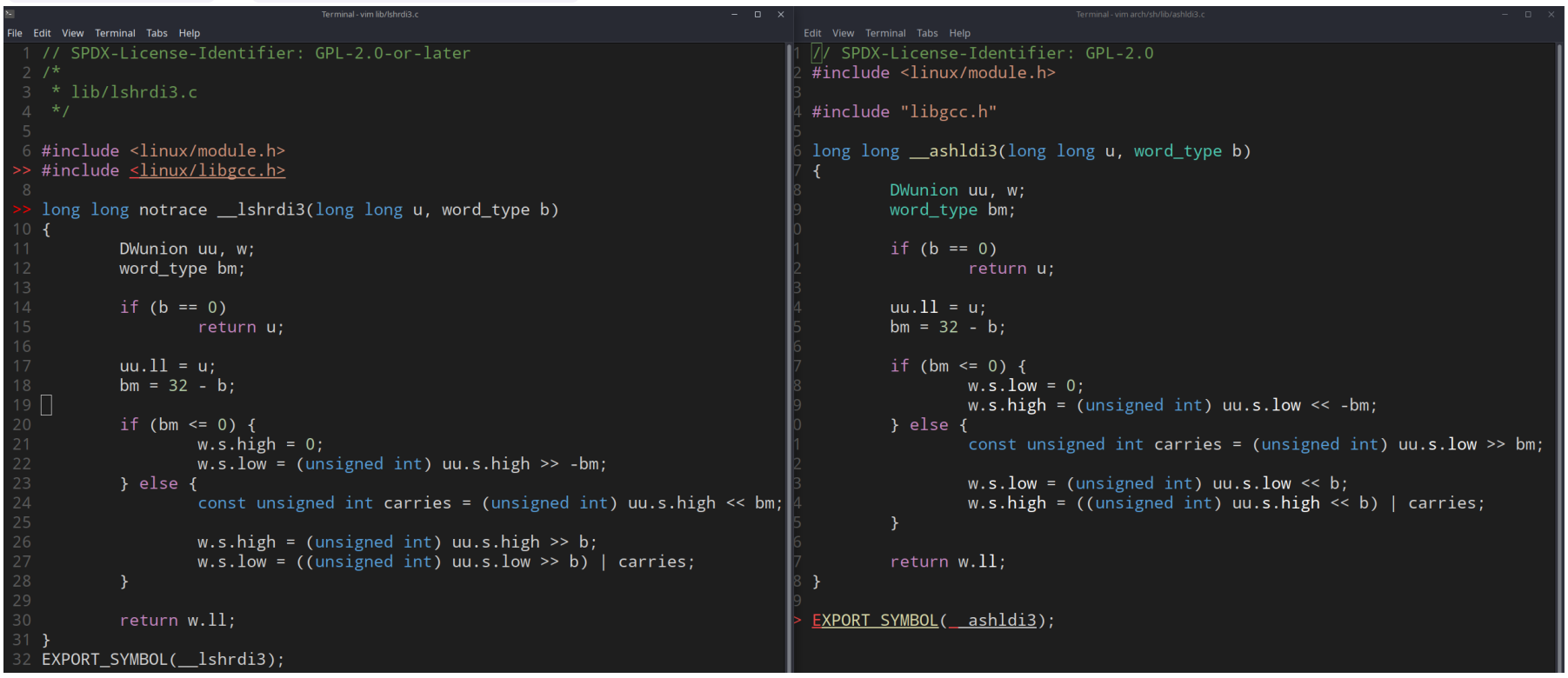
```
1,30955,1,3027
1,25101,1,3637
1,25166,1,3580
1,25210,1,1563
1,24336,1,4665
1,22820,1,9934
1,20602,1,528
1,20952,1,5260
1,3607,1,987
1,3897,1,7
...
```

这些相似为：

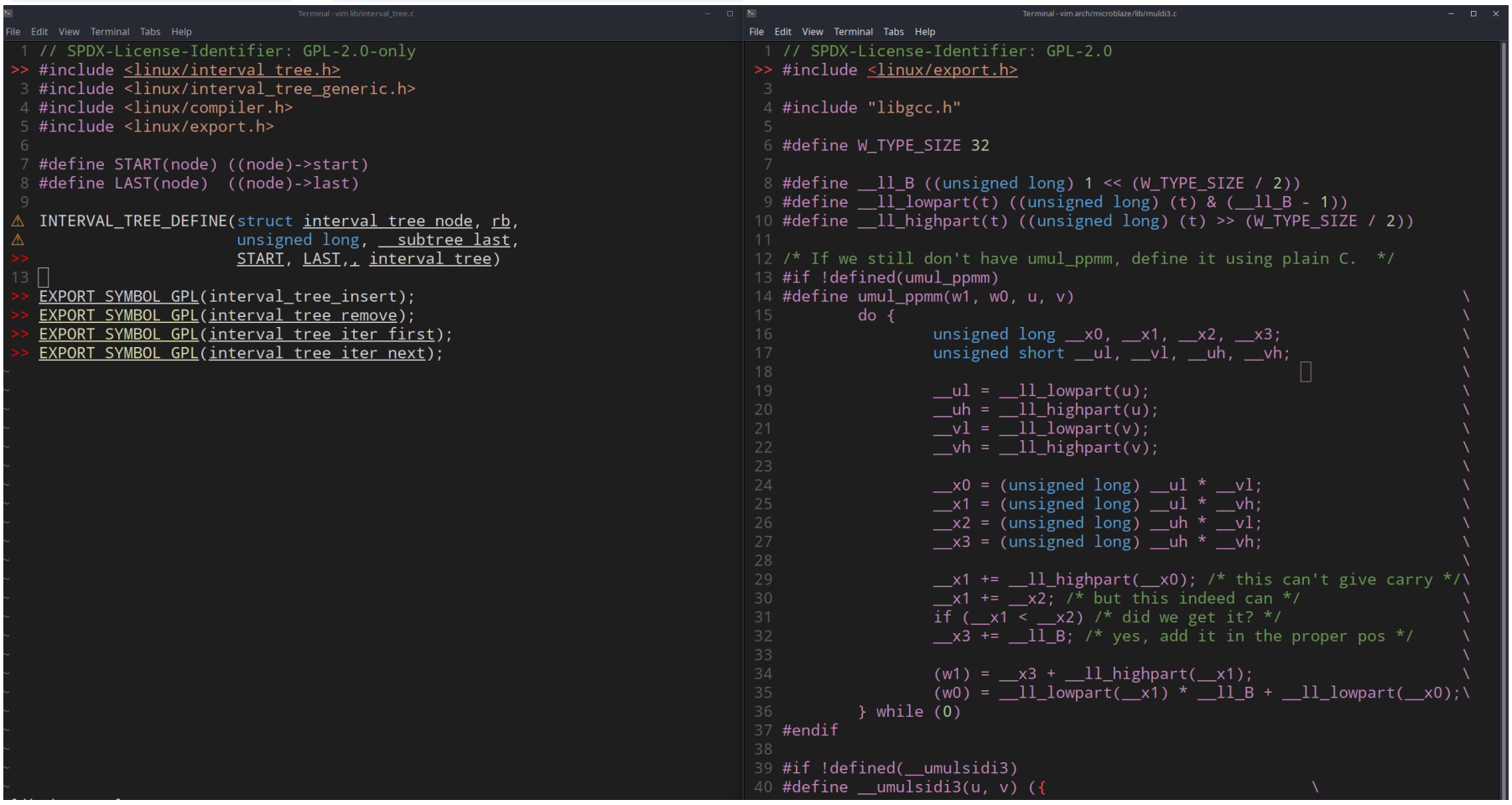
- `tools/testing/selftests/powerpc/vphn/vphn.c` 与 `arch/powerpc/platforms/pseries/vphn.c`
- `lib/fonts/font_8x16.c` 与 `arch/sparc/kernel/btext.c`：



- `lib/lshrdi3.c` 与 `arch/sh/lib/ashldi3.c`：



- `lib/interval_tree.c` 与 `arch/microblaze/lib/muldi3.c`：



- `fs/smbfs_common/cifs_md4.c` 与 `crypto/md4.c`
- `drivers/video/fbdev/omap2/omapfb/dss/dispc_coefs.c` 与 `drivers/gpu/drm/omapdrm/dss/dispc_coefs.c`

- `drivers/sh/pm_runtime.c` 与 `arch/arm/mach-omap1/pm_bus.c`：

Terminal - vim drivers/sh/pm_runtime.c

```
File  Edit  View  Terminal  Tabs  Help
 1 /*
 2  * Runtime PM support code
 3  *
 4  *  Copyright (C) 2009-2010 Magnus Damm
 5  *
 6  * This file is subject to the terms and conditions of the GNU General Publ
   c
 7  * License.  See the file "COPYING" in the main directory of this archive
 8  * for more details.
 9  */
10
>> #include <linux/init.h>
12 #include <linux/kernel.h>
13 #include <linux/io.h>
14 #include <linux/pm_runtime.h>
15 #include <linux/pm_domain.h>
16 #include <linux/pm_clock.h>
17 #include <linux/platform_device.h>
18 #include <linux/clk.h>
19 #include <linux/sh_clk.h>
20 #include <linux/bitmap.h>
21 #include <linux/slab.h>
22
>> static struct dev_pm_domain default_pm_domain = {
24         .ops = {
>>                 USE_PM_CLK_RUNTIME_OPS
26                 USE_PLATFORM_PM_SLEEP_OPS
27         },
28 };
29
>> static struct pm_clk_notifier_block platform_bus_notifier = {
31         .pm_domain = &default_pm_domain,
>>         .con_ids = { NULL, },
33 };
34
>> static int __init sh_pm_runtime_init(void)
36 {
37         pm_clk_add_notifier(&platform_bus_type, &platform_bus_notifier);
38         return 0;
39 }
```

Terminal - vim arch/arm/mach-omap1/pm_bus.c

```
File  Edit  View  Terminal  Tabs  Help
 1 /*
 2  * Runtime PM support code for OMAP1
 3  *
 4  * Author: Kevin Hilman, Deep Root Systems, LLC
 5  *
 6  * Copyright (C) 2010 Texas Instruments, Inc.
 7  *
 8  * This file is licensed under the terms of the GNU General Public
 9  * License version 2. This program is licensed "as is" without any
10  * warranty of any kind, whether express or implied.
11  */
>> #include <linux/init.h>
13 #include <linux/kernel.h>
14 #include <linux/io.h>
15 #include <linux/pm_runtime.h>
16 #include <linux/pm_clock.h>
17 #include <linux/platform_device.h>
18 #include <linux/mutex.h>
19 #include <linux/clk.h>
20 #include <linux/err.h>
21
22 #include "soc.h"
23
>> static struct dev_pm_domain default_pm_domain = {
25         .ops = {
>>                 USE_PM_CLK_RUNTIME_OPS
27                 USE_PLATFORM_PM_SLEEP_OPS
28         },
29 };
30
>> static struct pm_clk_notifier_block platform_bus_notifier = {
32         .pm_domain = &default_pm_domain,
>>         .con_ids = { "ick", "fck", NULL, },
34 };
35
>> static int __init omap1_pm_runtime_init(void)
37 {
38         if (!cpu_class_is_omap1())
39                 return -ENODEV;
40
```

- `drivers/ssb/driver_chipcommon_sflash.c` 与 `drivers/bcma/driver_chipcommon_sflash.c`
- `arch/sh/mm/mmap.c` 与 `arch/arm/mm/mmap.c`
- `arch/x86/boot/printf.c` 与 `arch/alpha/boot/stdio.c`

我们详细分析一下 `arch/x86/boot/printf.c` 与 `arch/alpha/boot/stdio.c`

可以看到他们确实是克隆关系：

Terminal - vim arch/x86/boot/printf.c

```
File  Edit  View  Terminal  Tabs  Help
 1 // SPDX-License-Identifier: GPL-2.0-only
 2 /* -*- linux-c -*- ------------------------------------------------------
 3  *
 4  *   Copyright (C) 1991, 1992 Linus Torvalds
 5  *   Copyright 2007 rPath, Inc. - All Rights Reserved
 6  *
 7  * ----------------------------------------------------------------------
   /
 8
 9 /*
10  * Oh, it's a waste of space, but oh-so-yummy for debugging.  This
11  * version of printf() does not include 64-bit support.  "Live with
12  * it."
13  *
14  */
15
>> #include "boot.h"
17
18 static int skip_atoi(const char **s)
19 {
20         int i = 0;
21
22         while (isdigit(**s))
23                 i = i * 10 + *((*s)++) - '0';
24         return i;
25 }
26
27 #define ZEROPAD 1               /* pad with zero */
28 #define SIGN    2               /* unsigned/signed long */
29 #define PLUS    4               /* show plus */
30 #define SPACE   8               /* space if plus */
31 #define LEFT    16              /* left justified */
32 #define SMALL   32              /* Must be 32 == 0x20 */
33 #define SPECIAL 64              /* 0x */
34
35 #define __do_div(n, base) ({ \
36 int __res; \
37 __res = ((unsigned long) n) % (unsigned) base; \
38 n = ((unsigned long) n) / (unsigned) base; \
39 __res; })
```

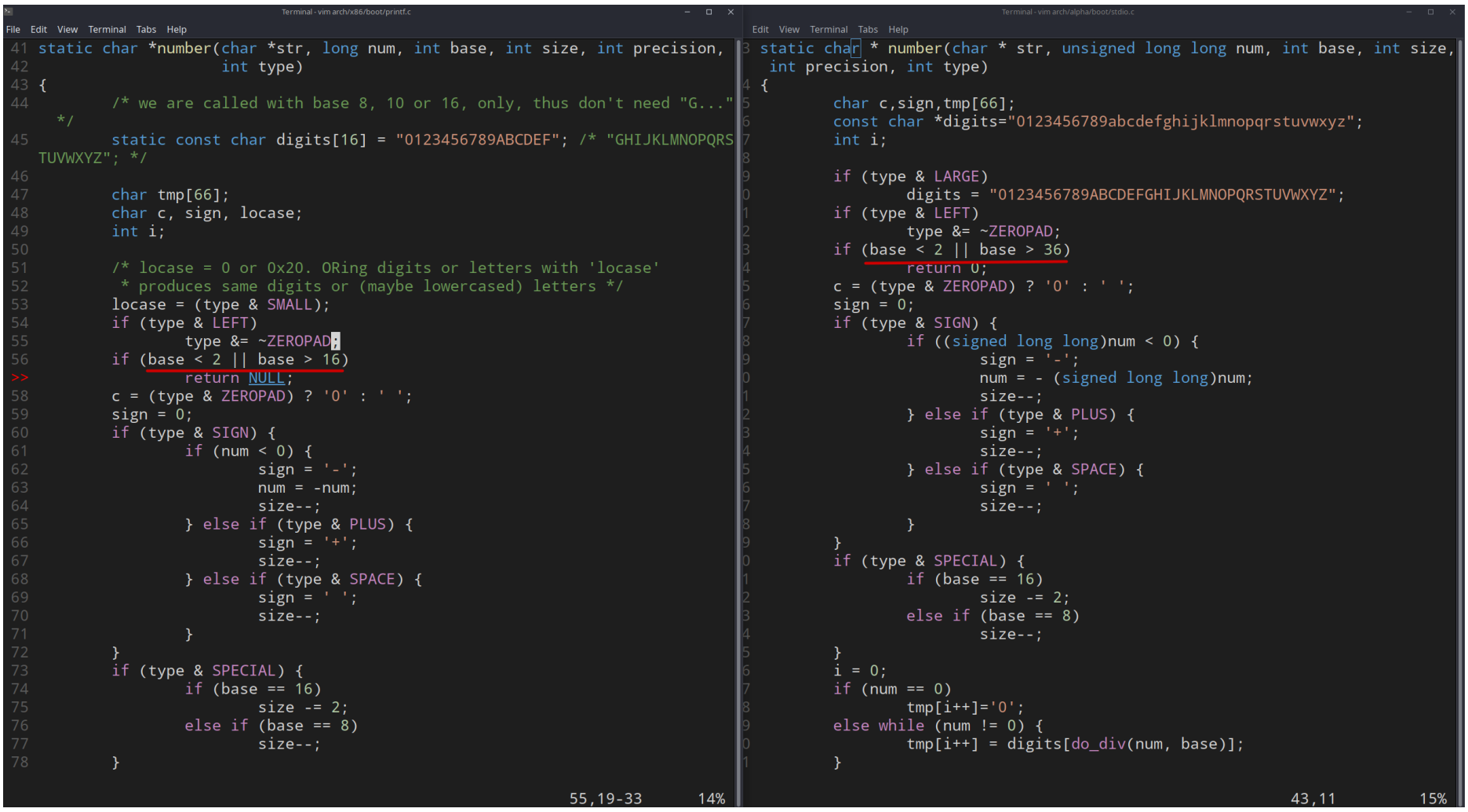
Terminal - vim arch/alpha/boot/stdio.c

```
File  Edit  View  Terminal  Tabs  Help
 8 size_t strnlen(const char * s, size_t count)
 9 {
10         const char *sc;
11
12         for (sc = s; count-- && *sc != '\0'; ++sc)
13                 /* nothing */;
14         return sc - s;
15 }
16
17 # define do_div(n, base) ({                                    \
18         unsigned int __base = (base);                          \
19         unsigned int __rem;                                    \
20         __rem = ((unsigned long long)(n)) % __base;            \
21         (n) = ((unsigned long long)(n)) / __base;              \
22         __rem;                                                 \
23 })
24
25
26 static int skip_atoi(const char **s)
27 {
28         int i, c;
29
30         for (i = 0; '0' <= (c = **s) && c <= '9'; ++*s)
31                 i = i*10 + c - '0';
32         return i;
33 }
34
35 #define ZEROPAD 1               /* pad with zero */
36 #define SIGN    2               /* unsigned/signed long */
37 #define PLUS    4               /* show plus */
38 #define SPACE   8               /* space if plus */
39 #define LEFT    16              /* left justified */
40 #define SPECIAL 32              /* 0x */
41 #define LARGE   64              /* use 'ABCDEF' instead of 'abcdef' */
42
43 static char * number(char * str, unsigned long long num, int base, int size,
    int precision, int type)
44 {
45         char c,sign,tmp[66];
46         const char *digits="0123456789abcdefghijklmnopqrstuvwxyz";
```

那么同样功能的函数，如果两边写法不一致，应该是以下情况之一：

1. 这个函数的实现特定于架构等上下文环境，所以实现思路相同，但具体细节有出入
2. 其中一方进行了冗余操作，例如
    - 指针使用前先检测是否为空，但这里指针不可能为空
    - 想要提升健全性而检测各种边界情况，但这里不可能出现部分情况
3. 其中一方有代码有BUG，比如另一方的操作并非冗余，而是这里缺少了一些操作；或者只是简单的与架构等特定环境无关的不一致

比如这里的出入是因为真的有36进制：



针对以上的一些分析，如果要用克隆检测来检查linux中一些某个地方修复而另一个地方忘记修复的漏洞，可以采取这样一些策略：

- 对于一些重要的漏洞修复，对修复的文件进行查重，看看有没有克隆的未被修复的文件
- 如果不去搜寻漏洞修复的话，可以对时间跨度较大的克隆对分析，那么较旧的一方（被遗忘的一方）可能会有BUG