**#################DEFACEMENTS#####################**

**#01#02#03#04#05#06#07#08#09#10#11#12#13#14#15#16#**

**#################################################**

**#..#..#..#..#..#..#..#..#..#..#..#..#..#..#..#D0#**

**#..#D0#D0#..#..#..#..#..#..#..#..#..#..#N0#..#D0#**

**#..#D0#..#..#..#N0#..#..#..#..#..#..#..#..#..#..#**

**#..#..#..#..#..#..#..#..#..#..#..#N1#..#..#..#..#**

**#..#..#D0#..#..#..#..#..#..#..#..#D0#..#..#..#..#**

**#..#..#D0#..#..#..#..#N1#..#..#..#..#..#..#..#..#**

**#..#..#..#..#..#..#..#..#N1#..#..#..#..#N0#..#..#**

**#################################################**

DX – element that appears in more than 3 defaces

NX – element that appears in most of 3 defaces

.. – element that does not have pair in any other daface (UN)

By observing burst of defaces of single notifier as depicted on draw, we are setting two presumption that are basis for detection algorithm.

**Presumption 1**. In burst of XX=16 defaces greatest possible number of signatures is AA=03.

**Presumption 2**. For elements of chain length up to BB=03 it is unknown if they are deface or *noise* elements, otherwise we presume they must be deface elements.

**Algorithm theory:**

In the expectation of notifier’s tendency for rare changes of one’s signature (presm. 1), we have opportunity to successfully eliminate noise elements and detect complete signatures. Algorithm returns set of *complete deface signatures*.

Basic idea goes behind *subgrouping operation* executed on three levels. DX, NX and UN subgrouping respectively. On each level there is a number of possibilities to include elements as deface elements (presuming rest of them to be noise), thus creating *families of subgroups*:

1. **DX subgrouping** -- there is none or one family, since all DX chains consist of deface elements by presm. 2.

**Bound condition 1.** : If DX chains subgrouping results in number of subgroups greater then AA, thus violating presm. 1, algorithm should finish immediately, returning empty set.

1. **NX subgrouping** – there is power set **𝒫** (NX chains) of families possible, since it is not certain if NX chain is consisted of deface or noise elements by presm. 2.
2. **UN subgrouping** – there is power set **𝒫** (UN) of families possible, since it is not certain if UN element is deface or noise elements by presm. 2.

Subgroup of certain family represents *deface subsignature*. Deface subsignature can be complete, namely *complete deface subsignature* or *deface signature*. **Each subgroup in family represents different deface subsignature.**

*Valid subgroups families* are families that obey presm. 1, taking into account that it is in advance known that number of potential deface signatures in family is equal or greater of number of deface subsignatures, i.e. number of subgroups in family.

Valid subgroups families on each level we are naming as *valid DX subgroups families, valid NX subgroups families, and valid UN subgroups families.* First of two are not *complete subgroups families*, as status of NX chains or NX chains and UN elements is unresolved.

**Filtering:**

Having all possible combinations of DX\*NX\*UN subgroups families it is necessary to filter out families they obey presm 1., thus limiting set of families to *valid complete subgroups families*.

**Signature search:**

Final step in algorithm would be to find all *deface signatures* that are present in **all** families. Found deface signatures are solutions to algorithm. By checking presence of deface signature in all families, with certainty we know it is notifier’s complete deface signature.

Executing brute-force DX\*NX\*UN subgrouping operation together with following steps would be computationally very expensive, therefore further optimizations are devised.

**Optimizing algorithm**

*Vertical optimization*

DX\*NX\*UN subgrouping can be depicted as *subgrouping tree* with nodes being families. Tree have a root in DX subgroup family, NX subgroups families in middle of tree and UN subgroup families at lower levels toward leafs.

Subgrouping of *non-valid* family will always produce family with same or greater amount of subgroups, thus making them non-valid family as well. Therefore, main idea around vertical optimization goes into cutting off subgrouping path in tree as soon subgrouping operation gives *non-valid family.* Vertically optimized subgrouping tree consists only out of valid DX, NX and UN subgroup families, thus creating valid subgrouping tree.

This way DX\*NX\*UN subgrouping operation produces only subset of all possible families, and it is avoided step of *filtering*.

DX subgrouping vertical optimization

It is |DX chains| long subgrouping operation, represented as root node in subgrouping tree.

**Optimization condition**: If DX chains subgrouping results in number of subgroups greater then AA, thus violating presm. 1, algorithm should finish immediately, returning empty set.

*Colouring*

DX subgrouping will give us single valid family (if any) of subgroups. This means that deface signatures found in final step of algorithm are certainly subset of DX family deface subsignatures. Having this in mind we focus algorithm on tracing subsignatures of DX subgroups. Colouring is tracing method during subgrouping operation and optimization of *signature search* step.

After subgrouping using only DX chains, we observe families of different NX chain inclusions. All the NX elements that are included as deface elements we denote as DDX. All excluded NX elements that we assume to be noise elements are denoted as NNX.

Each DX subgroup will be enumerated by giving *colour* to subgroup. Following NX subgrouping using subset of NX chains, will create in each step new set of subgroups. Final subgroups will contain DX, DDX, NNX and UN elements. *All subgroups that does not contain DDX elements, i.e. they were not affected by NX chain inclusion operations will inherit colour from subgroup they are derived from.*

Also note, that on the way forward to build subgroups families path certain colours can vanish as it is possible that complete coloured subgroup comes under NX chain inclusion. Colour that vanish in valid subgrouping tree is no longer traced as there is valid subgroup family that does not contain traced DX deface subsignature.

NX subgrouping vertical optimization

We observe combinations of NX element chains in sense to include or exclude chains as deface elements. All power set inclusion combinations are counted in following way (given example of 3 NX chains):

**{}**

**{0}**

**{1}**

**{2}**

**{01}**

**{02}**

**{12}**

**{123}**

**Optimization condition:** If during NX chains subgrouping of certain power set combination results in number of subgroups greater then AA, thus violating presm. 1, combination chain is not further subgrouped. Having chain index I on position P subgrouping violating presm. 1, subnumber[0-P] is increased by one, and minimum next number with given subnumber, greater than current number is used to proceed building valid families in tree.

Horizontal optimization

Horizontal optimization is optimization that affects subgrouping operation only on one level. It is different for NX I UN level, where vertical is same for DX and NX levels.

NX subgrouping horizontal optimization

Certain combination of chains *c1cc2c3…cn* is included one by one in subgrouping operation until we have final family. With each chain inclusion new set of subgroups is born. Colour is forwarded from DX subgroup family trough intermediate set of subgroups until final family.

Let’s assume there is certain subset1(c1, c2, …, cn) of NX chains that does not affect current coloured subgroups in subgrouping operation on the way of building final family. That means that all the rest subset2 = *c1cc2c3…cn -* subset1(c1, c2, …, cn) chains will change **some** coloured subgroup on the way of building final family.

**Therefore we conclude that final subgroup family F1 after *c1cc2c3…cn* chain subgrouping operation will have equal coloured subgroups as final subgroup family F2 after subset2 chain subgrouping operation.** **F1 may have equal or greater number of subgroups (|F1|>=|F2|).**

**Horizontal subgrouping operation is possible as it is not needed to trace F1 family as it will result with same coloured subgroups as F2 family or F1 family will be non-valid.**

Note that chain that creates F2 family is always shorter then F1 family chain, thus being first counted by counting algorithm described above.

Counting algorithm will skip any chain inclusion that results in same coloured subgroup as previous step chain inclusion. This way algorithm will count all chains inclusions combinations that are necessary and sufficient to be counted.

All families that are not generated as result of this optimization are called *virtual NX familes*.

UN subgrouping horizontal optimization

All defaces consist of DX elements (where it could be DX={}), and any combination of NX and UN elements (NX, UN, [NX, UN], []).

All defaces consist of DX, NX and UN elements where it could be **(DX, NX, UN) = ({}, {}, {})**.

We focus our attention to valid NX subgroup just before UN subgrouping operation. Let’s observe single subsignature group. All the NX elements that are included as deface elements we denote as DDX. All excluded NX elements that we assume to be noise elements are denoted as NNX. So all defaces now have equal subsignatures denoted as DX+DDX, they also may consist of different NNX set. Only elements left unsettled in this moment are unpaired elements (UN), if exists. Note that it could be **(DDX, NNX) = ({}, {})**.

All the UN elements that are included as deface elements we denote as DUN. All excluded UN elements that we assume to be noise elements are denoted as UUN. Note that it could be **(DUN, UUN) = ({}, {})**.

**Comment on colouring:**

Developing subgroups families path in tree on UN level with inclusions on subgroups with subsignatures **DX+DDX, DDX≠{}** is useless as it would be discarded as complete signature because non-existence of subsignature on DX level.

As colouring would give all subgroups in family with DDX={}, it is needed to proceed with coloured subgroups only operation on UN level. (ovo je isto optimizacija na koju treba obratiti paznju)

Subgrouping operation on UN level is necessary for **all** *valid NX families* and all *valid virtual NX families*.

*Valid NX families*

If UN element do not exists in some of defaces of certain subgroup (DDX={}, DUN={}), then all valid families in path derived from that family will have DX+DDX+DUN=DX as *complete deface signature*.

If UN element exists in **all** of defaces of certain subgroup S (DDX={}, UN≠{} for all defaces) of some family **F**, we use following conclusion to prove existence/non-existence of D*X complete deface signature* in all families derived from family **F**.

Since all UN elements are different, inclusion of any number (>0) of elements in **all** defaces of subgroup **S** would create minimum of |F|-1+|S| subsignatures. If number of subsignatures obeys persm 1. (AA>=|F|-1+|S|), thus creating valid subgroup family, then search for *DX complete deface signature* is stopped. There exists at least one family with no *DX complete deface signature,* one where there is no subgrouping operation done to any other subgroup except **S**.

If **AA<|F|-1+|S|**, violating presm 1., thus creating non-valid subgroup family, then **all** valid subgroups families derived from **S** (if any valid exists), would certainty not have included UN elements of all defaces. Instead, there would exists deface in subgroup S having UN elements not included, thus creating new subgroup with DDX={}, DUN={} and DX+DDX+DUN=DX as complete deface signature.

*Valid virtual NX families*

Each virtual NX family F1 have corresponding valid NX family F2. They both have equal coloured subgroups and |F1|>=|F2|. If F1 is not valid (|F1|>=AA) then is not relevant for observation, and all colours of F2 should be kept after processing of F2. If F1 is valid then colouring result of UN level subgrouping operation would:

|  |  |  |  |
| --- | --- | --- | --- |
|  | F2 family test | F1 family consequence | Colour results for F1 and F2 |
| **NON-VALID VIRTUAL FAMILY F1** | Valid F2 family. Colour passes to level 3 subgrouping. | Non-valid F1 family. Family is discarded from subgrouping tree. | Colour is passing. |
| **VALID VIRTUAL FAMILY F1** | Colour passes UN subgrouping test | Since |F1|>=|F2| and AA<|F2|-1+|S|, then  |F1|-1+|S|>AA.  Colour will pass on test. | Colour is passing. |
| Colour fails UN subgrouping test | Since |F1|>=|F2| and AA>=|F2|-1+|S|, we can be sure if colour would pass. | Colour is vanishing. |

Currently we observe valid number of DX+DDX subsignature groups.

All UN inclusion combinations that create number of subgroups that obey pres.1 are observed (valid UN combinations).

By looking all UN inclusion subgroups we focus our attention to defaces belonging to certain DX+DDX subgroup. Those defaces now have included different DUN elements (UN elements assumed to be deface elements). They have also included different UUN elements, UN elements assumed to be noise elements. In fact, there are defaces having included any of combination {DUN, UUN, [DUN, UUN], []}.

We need to prove that under all valid UN combinations, there always exist at least one deface in observed DX+DDX subgroup, without any DUN element. This leads to conclusion that in all valid UN combinations DX+DDX is *complete deface signature*. That leads to conclusion that under observed NX inclusion DX+DDX is complete deface signature.

Proof: Let’s observe UN elements of given DX+DDX subgroup before inclusion operation.

There may be defaces in subgroup having no UN element, or **all** defaces having UN elements.

1. If there exists deface having no UN element, then under all valid UN inclusion combinations observed deface is *complete deface signature*. If there is no single valid UN inclusion than observed deface is complete deface signature in only possible subgrouping – NX only inclusion.
2. If **all** defaces have UN elements then we use following rule to check that under all valid UN inclusion combinations observed deface is *complete deface signature.*

IF **(subgroups in NX only inclusion – 1 + number of members in observed DX+DDX)** is greater than AA, that means that all valid UN inclusions contain DX+DDX defaces that always have at least one deface without DUN.