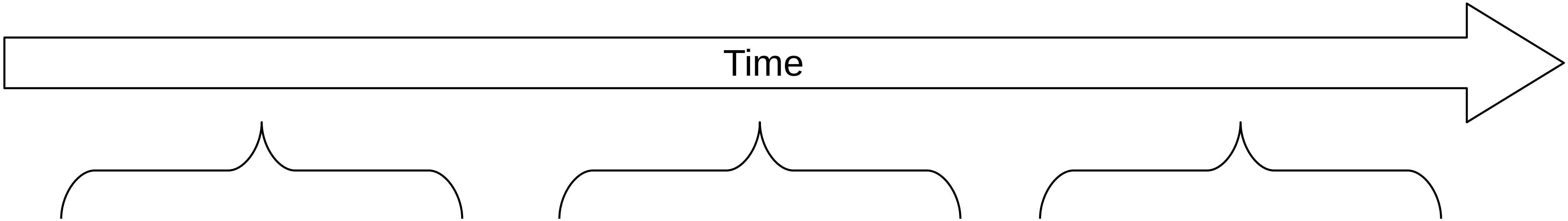
Towards Automated Risk Analysis of “One-day” vulnerabilities

Clément Elbaz – Univ Rennes, Inria, CNRS, IRISA
Louis Rilling - DGA
Christine Morin - Univ Rennes, Inria, CNRS, IRISA

Inria – MYRIADS team

Once upon a time: the journey of a vulnerability



Time

Zero-day vulnerability

- Confidential
- Used sparingly on high-value targets

“One-day” vulnerability

- Disclosure is a dangerous time!

Public vulnerability

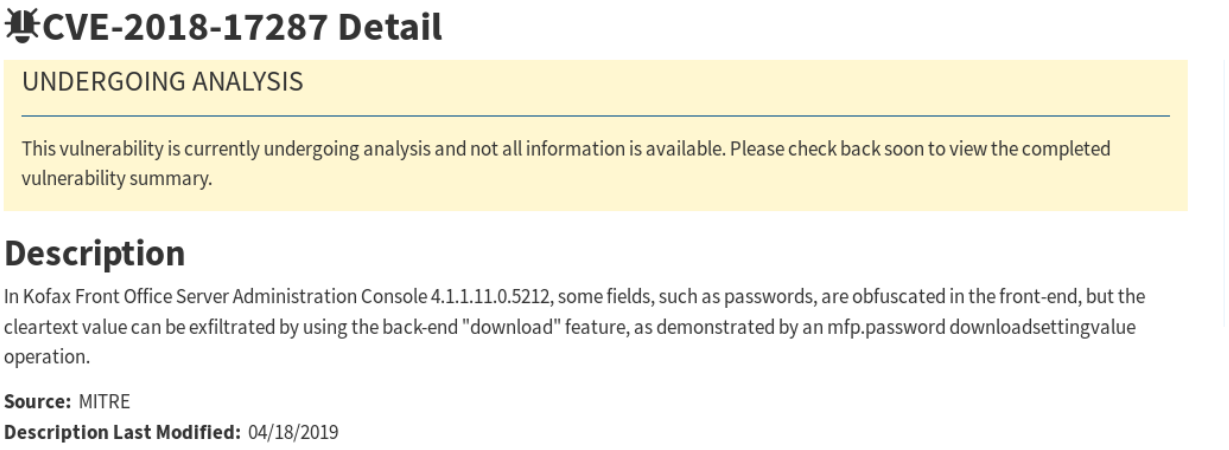
- Mitigated using standard security practices
  - Software updates, IDS

Disclosure is a dangerous time

- Usage of vulnerabilities increase as high as five orders of magnitude once disclosed
  - L. Bilge and T. Dumitraş, “Before We Knew It: An Empirical Study of Zero-day Attacks in the Real World,” in Proceedings of the 2012 ACM Conference on Computer and Communications Security, ser. CCS ’12. New York, NY, USA: ACM, 2012, pp. 833–844.
- Software patches may be available, but adoption is not widespread yet
- Vulnerability is not understood well yet
  - Metadata is either missing or sparse
  - No IDS signature rules yet

Metadata is late

- On disclosure day: an ID, a description, a link



CVE-2018-17287 Detail

UNDERGOING ANALYSIS

This vulnerability is currently undergoing analysis and not all information is available. Please check back soon to view the completed vulnerability summary.

Description

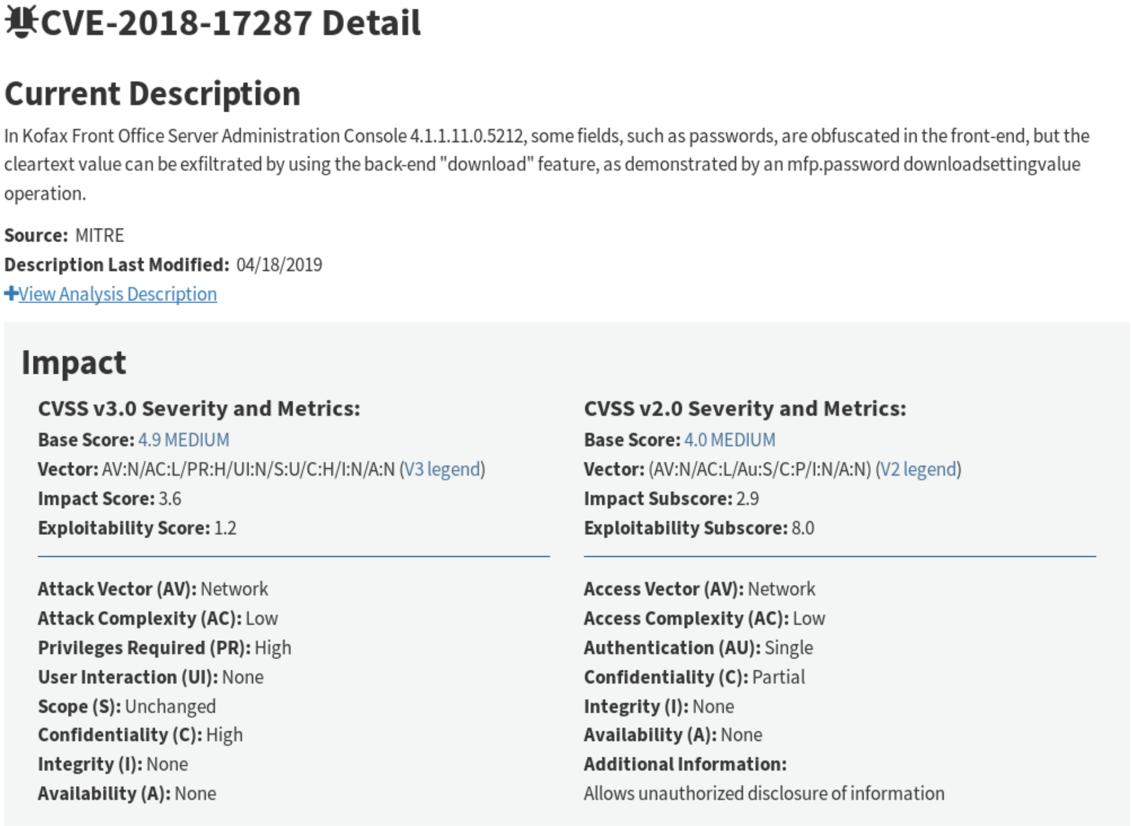
In Kofax Front Office Server Administration Console 4.1.1.11.0.5212, some fields, such as passwords, are obfuscated in the front-end, but the cleartext value can be exfiltrated by using the back-end "download" feature, as demonstrated by an mfp.password downloadsettingvalue operation.

Source: MITRE
Description Last Modified: 04/18/2019

https://nvd.nist.gov/vuln/detail/CVE-2018-17287 on **04/18/2019**

Metadata is late

- Analysis comes later
  - At best after one day, at worst after **six days**!



CVE-2018-17287 Detail

Current Description

In Kofax Front Office Server Administration Console 4.1.1.11.0.5212, some fields, such as passwords, are obfuscated in the front-end, but the cleartext value can be exfiltrated by using the back-end "download" feature, as demonstrated by an mfp.password downloadsettingvalue operation.

Source: MITRE
Description Last Modified: 04/18/2019
+View Analysis Description

Impact

CVSS v3.0 Severity and Metrics:
Base Score: 4.9 MEDIUM
Vector: AV:N/AC:L/PR:H/UI:N/S:U/C:H/I:N/A:N (V3 legend)
Impact Score: 3.6
Exploitability Score: 1.2

Attack Vector (AV): Network
Attack Complexity (AC): Low
Privileges Required (PR): High
User Interaction (UI): None
Scope (S): Unchanged
Confidentiality (C): High
Integrity (I): None
Availability (A): None

CVSS v2.0 Severity and Metrics:
Base Score: 4.0 MEDIUM
Vector: (AV:N/AC:L/Au:S/C:P/I:N/A:N) (V2 legend)
Impact Subscore: 2.9
Exploitability Subscore: 8.0

Access Vector (AV): Network
Access Complexity (AC): Low
Authentication (AU): Single
Confidentiality (C): Partial
Integrity (I): None
Availability (A): None
Additional Information:
Allows unauthorized disclosure of information

https://nvd.nist.gov/vuln/detail/CVE-2018-17287 on **04/24/2019**

In order to reliably analyze one-day vulnerabilities, we have to rely on their text description only.

Goals

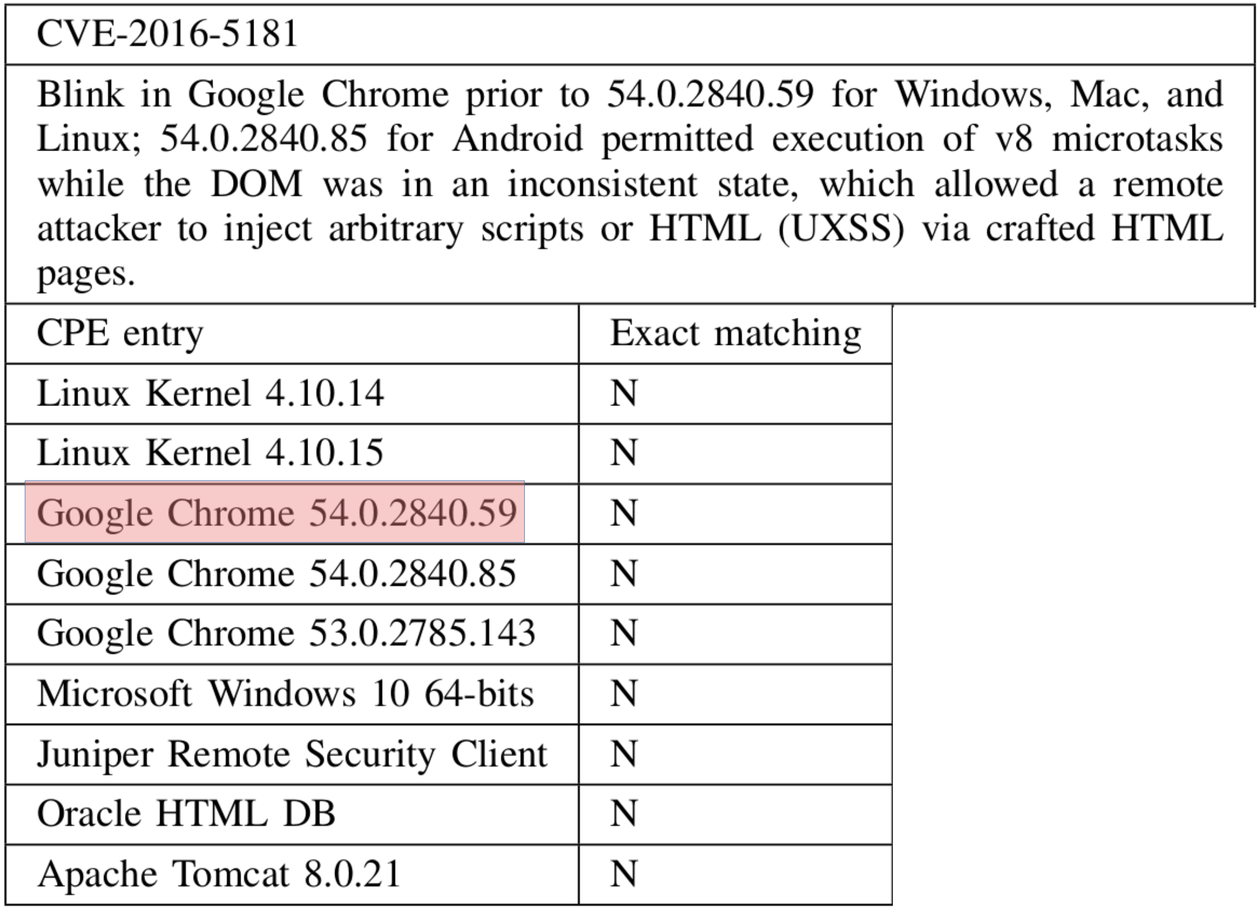
- Automated threat assessment of one-day vulnerabilities
  - In the context of a specific information system
- A first step: deducing the affected software from the text description

Extracting the affected software from the description

- A first contribution: automated mapping of a vulnerability to CPE dictionary entries
  - The CPE dictionary references every software ever afflicted by a vulnerability
- Explainability and simplicity are paramount for security
- Still, we want reasonable accuracy

Mapping techniques: from the most simple to the most accurate

- Exact pattern matching
  - If the entry is spelled out in the description, there is a match
  - Fail in practice: most descriptions do not spell exact CPE entries explicitly



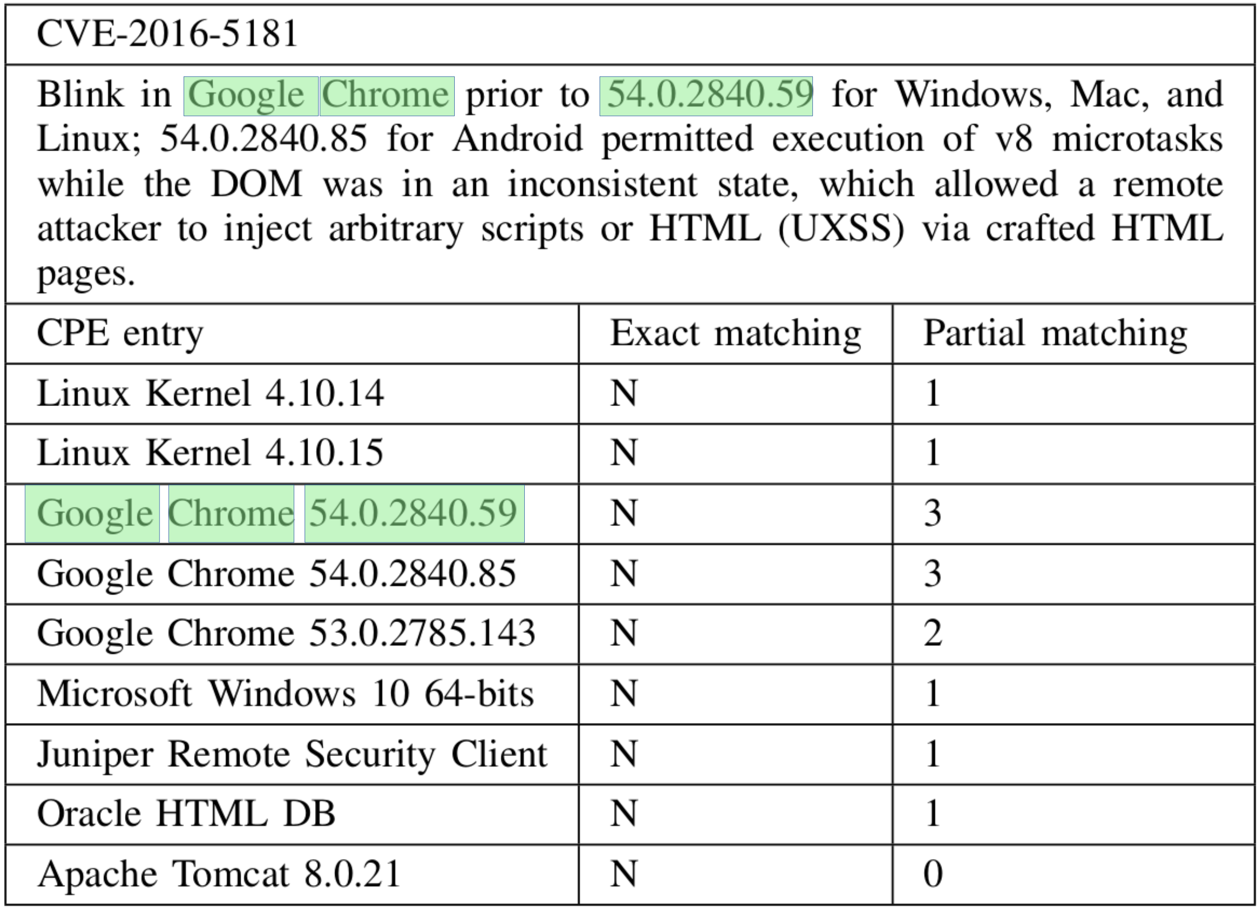
CVE-2016-5181

Blink in Google Chrome prior to 54.0.2840.59 for Windows, Mac, and Linux; 54.0.2840.85 for Android permitted execution of v8 microtasks while the DOM was in an inconsistent state, which allowed a remote attacker to inject arbitrary scripts or HTML (UXSS) via crafted HTML pages.

CPE entry	Exact matching
Linux Kernel 4.10.14	N
Linux Kernel 4.10.15	N
Google Chrome 54.0.2840.59	N
Google Chrome 54.0.2840.85	N
Google Chrome 53.0.2785.143	N
Microsoft Windows 10 64-bits	N
Juniper Remote Security Client	N
Oracle HTML DB	N
Apache Tomcat 8.0.21	N

Mapping techniques: from the most simple to the most accurate

- Partial pattern matching
  - Each description and CPE entries are tokenized into individual words
  - Every common word increment a score
  - Too many false positives to be usable



CVE-2016-5181

Blink in Google Chrome prior to 54.0.2840.59 for Windows, Mac, and Linux; 54.0.2840.85 for Android permitted execution of v8 microtasks while the DOM was in an inconsistent state, which allowed a remote attacker to inject arbitrary scripts or HTML (UXSS) via crafted HTML pages.

CPE entry	Exact matching	Partial matching
Linux Kernel 4.10.14	N	1
Linux Kernel 4.10.15	N	1
Google Chrome 54.0.2840.59	N	3
Google Chrome 54.0.2840.85	N	3
Google Chrome 53.0.2785.143	N	2
Microsoft Windows 10 64-bits	N	1
Juniper Remote Security Client	N	1
Oracle HTML DB	N	1
Apache Tomcat 8.0.21	N	0

Mapping techniques: from the most simple to the most accurate

- Weighted partial pattern matching
  - Instead of incrementing by a fixed amount, add the TF-IDF value of the matched word
- Term frequency – Inverse Document Frequency

$$\mathrm{TFIDF}(t, d, D) = \mathrm{TF}(t, d) \times \mathrm{IDF}(t, D)$$

  - ***t*** is a word, ***d*** is a document belong to a corpus ***D***
  - ***TF(t, d)*** is the number of occurences of a word **t** in a document **d**
  - **IDF(t, D)** = $\log \frac{|D|}{|d \in D : t \in d|}$

Conclusion and future work

- Preliminary results available
  - Current accuracy at 66%: promising but low for some applications
  - Work in progress: some low hanging fruits on the roadmap
  - See poster and paper for details
- Towards automated threat assessment of one-day vulnerabilities in the context of a given IS
- Source code available
  - https://gitlab.inria.fr/celbaz/firres_ressi

Backup slides

Results

- Evaluation dataset
  - CVE corpus for the year 2016 : 8068 vulnerabilities
  - CPE dictionary in version 2.3 : 17631 pieces of software (124681 entries with unique versions)
- Manual evaluation of 229 vulnerabilities
  - Answering the following question
    - *“Does the top 3 mapped propositions includes at least one actually afflicted software ?”*
  - 151 correctly classified vulnerability
  - A **66 % success rate**