Tell me you fixed it

Evaluating vulnerability notifications via quarantine networks

Cetin, Orcun; Ganan, Carlos; Altena, Lisette; Tajalizadehkhoob, Samaneh; Van Eeten, Michel

DOI
10.1109/EuroSP.2019.00032

Publication date
2019

Document Version
Final published version

Published in
Proceedings - 4th IEEE European Symposium on Security and Privacy, EURO S and P 2019

2019 IEEE European Symposium on Security and Privacy (EuroS&P)

Tell Me You Fixed It: Evaluating Vulnerability Notifications via Quarantine Networks

Orçun Çetin, Carlos Gañán, Lisette Altena, Samaneh Tajalizadehkhoob, Michel van Eeten
Delft University of Technology, the Netherlands
Email: {f.o.cetin, c.hernandezganan, e.m.altena, s.t.tajalizadehkhoob, m.j.g.vaneeten}@tudelft.nl

***Abstract*—Mechanisms for large-scale vulnerability notifications have been confronted with disappointing remediation rates. It has proven difficult to reach the relevant party and, once reached, to incentivize them to act. We present the first empirical study of a potentially more effective mechanism: quarantining the vulnerable resource until it is remediated. We have measured the remediation rates achieved by a medium-sized ISP for $1,688$ retail customers running open DNS resolvers or Multicast DNS services. These servers can be abused in UDP-based amplification attacks. We assess the effectiveness of quarantining by comparing remediation with two other groups: one group which was notified but not quarantined and another group where no action was taken. We find very high remediation rates for the quarantined users, 87%, even though they can self-release from the quarantine environment. Of those who received the email-only notification, 76% remediated. Surprisingly, over half of the customers who were not notified at all also remediated, though this is tied to the fact that many observations of vulnerable servers are transient. All in all, quarantining appears more effective than other notification and remediation mechanisms, but it is also clear that it can not be deployed as a general solution for Internet-wide notifications.**

I. Introduction

Our ability to undertake large-scale vulnerability discovery has grown immensely, providing a wealth of data on vulnerable resources to help those responsible for the affected resources. Notification and remediation, however, has proven to be much harder. Randomized controlled experiments with different notification mechanisms have found remediation rates that typically range from modest to abysmal. These low rates persisted across disclosures via email, national CSIRTs (Computer Security Incident Response Teams), social networks and even phone calls [1]–[4].

There are varying explanations for the disappointing remediation rates. Most experiments used email. Because it scales reasonably well, this is still the dominant channel for notifications. Reachability has proven to be a key problem, however. Notifications are sent to addresses that are RFC-specified or harvested from WHOIS records. Delivery is severely hampered by non-existing email addresses and poorly-configured spam filters. When messages are actually received and read, there is often no follow-up action. These problems are not specific to email. Even more manual methods for notifications, such as postal mail or phone calls, have the same issue [3]. The lack of follow-up actions points to problems with trust, technical competency and lack of incentives for remediation.

One would expect that the incentive problem would be even worse for vulnerabilities that threaten third parties rather than the party responsible for the vulnerable resource. Think of NTP servers that can be abused in UDP-based amplification DDoS attacks against any target on the Internet. They are rarely, if ever, used against the party responsible for the vulnerable server itself. Remarkably, though, a 2013 campaign of researchers and the security community managed to reduce the number of vulnerable NTP amplifiers by more than 92% in three months [5].

This stand-out success has been difficult to interpret, partly because it was not a randomized controlled experiment. A high-profile campaign that did use an experimental design, was Heartbleed [6]. It also found a relatively high overall remediation rate of around 60% over the course of a month. While these examples provide inspiring counterpoints to the studies with disappointing remediation rates, these high-profile campaigns do not seem suitable templates for large-scale vulnerability notifications.

All in all, prior work on notifications has observed an alarming and increasing discrepancy between the community's ability to gather vulnerability data and its ability to make this information useful for preventing future abuse. In this paper, we empirically explore the effectiveness of an alternative mechanism for vulnerability notification and remediation: quarantining the vulnerable resource in a so-called walled garden environment. We compare this to the current default approach: email notifications. Walled gardens tackle both challenges identified in previous studies. First, it provides a much more robust mechanism to notify the responsible party, as Internet access is restricted and a landing page informs the party responsible for the vulnerable device of the reason why the connection is quarantined. In other words, it is almost impossible to overlook the notification. Second, the mechanism increases the incentive to remediate, as release from the walled garden is conditional on remediation. Prior studies has found that quarantining was effective in cleaning malware infections in ISP networks [7], [8]. Its effectiveness in remediating vulnerable resources, however, has never been studied before.

We study a walled garden implementation for vulnerable resources at a medium-sized Internet Service Provider (ISP). We measured remediation rates for $1,688$ retail customers with servers running open DNS resolvers or Multicast DNS services, which can be abused in amplification DDoS attacks.

DOI 10.1109/EuroSP.2019.00032

We assess the effectiveness of quarantining by comparing remediation with two other groups: one group which was only notified by an email but not quarantined and another group where no action was taken.

In short, we make the following contributions:

- We present the first empirical study of the remediation effectiveness of quarantining vulnerable resources. Even though customers can self-release from the quarantine environment without actually remediating the problem, we find very high remediation rates of around 87%. Of those who received only the email notification, around 75% remediated.
- We find a remarkably high remediation rate in the control group: around half of all customers remediate. This high rate reflects actual remediation actions, but also the fact that a significant portion of the observations of vulnerable devices are transient. These observations have typically been omitted from prior studies, which might explain the low remediation rates reported in those papers. This might reflect selection bias.
- We analyze communications between notified customers and the ISP to assess challenges in remediation. We find out that 16% of the notified users were unwilling to remediate because they did not want to change the way they use their device. Around 11% of the notified users complained about the disruptiveness of being quarantined in the walled garden.

Notwithstanding the potential advantages of walled garden solutions, we want to emphasize that quarantining vulnerable resources is not a silver bullet. Quarantining by network operators is only feasible under certain scenarios. There are also downsides in terms of cost and customer pushback. We will discuss these in the course of the paper. We do argue, however, that there is an urgent need to find more effective notification and remediation mechanisms. This puts a premium on examining solutions for which no prior empirical studies exist.

This paper is structured as follows. Section II explains the unique natural experiment that was inadvertently conducted by a European ISP. Section III describes the data collection mechanism. Section IV evaluates the effectiveness of walled garden and email notification mechanisms compared to natural remediation. Section V presents key insights gathered from communications. Section VI evaluates prior work and explains how this is related to ours. We outline ethical considerations and limitations of the study in Section VII and VIII and conclude the study in Section IX.

II. Vulnerability Notification Experiment

For this study, we collaborated closely with a European ISP which operates in various markets. Here, we will focus on its retail broadband services, which have around 2 million customers. A few years ago, the ISP implemented its first version of a walled garden solution to deal with malware infections among its retail customers. More recently, the ISP started allocating spare capacity in the walled garden environment to undertake notification and remediation for users with devices that are vulnerable to UDP-based amplification attacks, as identified in Shadowserver scans for such amplification factors. The ISP only does these notifications on a fixed day each week. This setup provides a natural experiment, as the assignment of customers to one of three groups (quarantine, email, no action) is more or less random.

A. Walled garden notifications

In the early days of the Internet, the concept of a walled garden referred to a closed environment that restricted the content and services that users could access. Nowadays, a walled garden primarily refers to a security best practice in botnet mitigation [9], as described in RFC6561 [10]. It is a method to notify affected users about a security problem and quarantine their connection to prevent the infected machine from being abused by miscreants.

The ISP with which we collaborated has adopted a so-called strict implementation of a walled garden. This means that the quarantine network redirects all web browsing activity to a landing page, except for a small set of white-listed sites. The landing page explains the problem and provides guidance on resolving it (see Appendix A). The advantage of this notification mechanism, compared to email, postal and phone notifications, is that it is much less likely to be overlooked or ignored. At the same time that the connection is quarantined, the ISP sends an email to the customers with the same information as the landing page. Thus, users don't need to be at their home to understand that their Internet connection has been quarantined by the ISP.

There are three ways the customer can get out of the walled garden. First and foremost, customers can release themselves from the quarantine environment via a button underneath a form for reporting on what action was taken. The self-release option is revoked after two subsequent quarantine events in the same month, to avoid customers using this route to restore their connection without making an effort at remediation. The second way out is when the ISP's abuse staff releases the customer's connection. Customers might end up in assisted release because they no longer have the self-release option or because they have contacted the ISP for help. Quarantined customers can contact abuse desk members via email and a walled-garden form. The third way of being released is when the expiration date passes. After 30 days, a customer is automatically released, even if they have not contacted the ISP.

B. Email notifications

The walled garden has a limited capacity. When all slots are taken, but the ISP still wants to notify and remediate, it can send an email notification to the mail address that it has on record as the primary contact for that customer. For some customers, the ISP's mail service is the primary contact point. For other users, it does not have full visibility into the delivery

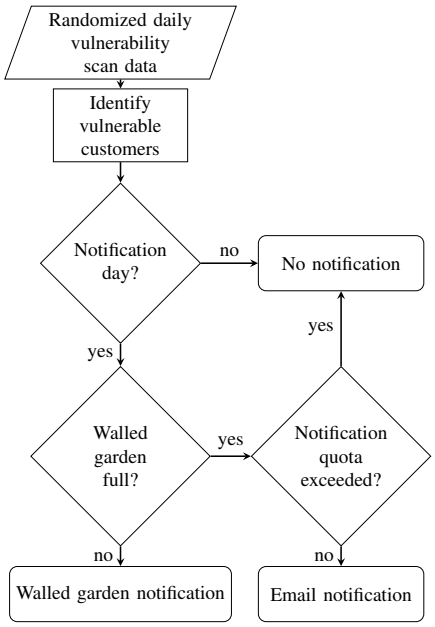


Fig. 1: Vulnerability notification flowchart

success of the message. That being said, these are email addresses that were supplied by the customers themselves, so the odds of success are a lot higher than mailing RFC-specified addresses or generic WHOIS contact points. The message contains the same information as the walled garden's landing page, plus an email address to contact in case of questions or problems while remediating the vulnerability.

C. Notification process and assignment mechanisms

On a daily basis, the ISP receives vulnerability scan data from third parties, most notably Shadowserver, specifying a list of vulnerable IP addresses in the network. IP addresses show up in the daily vulnerability scan data in a random order. Because of time constraints of the abuse department, the ISP notifies owners of these resources only once per week, with different vulnerabilities being assigned different weekdays. For mDNS and open resolver notifications are made every Thursday, using the IP addresses from Wednesday's reports. This arbitrary policy and randomized list of IP addresses in the vulnerability scan data create a natural experiment: a *de facto* random assignment to being notified or not notified, assuming that there is no systematic difference between customers that show up in Wednesday's reports versus the reports from, say, Tuesday or Thursday.

The next step contains a random assignment between the two treatment conditions: walled garden and email-only notifications. The ISP's walled garden can fit up to 100 customers at any time. Many of the slots are taken for higher priority issues, such as malware infections. The remaining slots are dedicated to a random batch of customers selected from Wednesday's Shadowserver report, without any prior inspection of the IP addresses in the report. When full capacity is reached, the remaining customers are notified via email, until also for that treatment a quota is reached. The quota is a bit fuzzy and depends on the available resources (e.g., abuse department staff, number of open tickets, etc.). If the walled garden capacity and the email notification quota are both exceeded, then the remaining vulnerable customers are not notified. Figure 1 shows a flowchart of the treatment assignment process. A direct consequence of only notifying once per week is a higher amount of vulnerable customers that do not receive the treatment compared to the ones that are not notified. This imbalance will increase the power of the natural experiment even though the groups are asymmetric in size.

Given this notification process, the treatment assignment is independent of the characteristics of the vulnerable population. When after a notification a customer machine shows up again in the Shadowserver reports, then the assignment process may result in a subsequent treatment. Also, customers may have other vulnerabilities on their machine and they may also receive notifications for these issues via a different procedure, delivered on other weekdays. This may also impact the remediation of mDNS and Open resolver vulnerabilities. In our statistical analysis, we use an instrumental variable to account for this effect on the vulnerability remediation.

D. Other walled garden notifications

As this experiment was conducted in a real-world setting, we also had to take into account that the ISP sent out notifications for other security and vulnerability issues that were not part of the experiment. Checking the ISP logs, we found out that 231 users in our study did in fact receive another walled garden notification (16% of the users in the control group and 8% of the users in the treatment groups, see Table III). The bulk of these notifications (95%) were for NetBIOS. Like mDNS and Open resolver, NetBIOS can be abused in amplification attacks. Rather than removing these users from the study, we decided to keep them in and use this opportunity to study the impact of other notification processes. Most real-world randomized controlled notification experiments are likely impacted by unobserved 'parallel' notifications processes. In those cases, the researchers typically have no data on this. In our study, we did have the data, so we were in a position to identify just how these 'other notifications' impacted the results.

III. Data Collection

To assess the notification and remediation success of the walled garden solution, we correlate three different datasets collected by the ISP: (*i*) Daily scan results on the presence of vulnerable amplifiers in the ISP's network, provided to the ISP by the Shadowserver Foundation; (*ii*) ISP logs that capture the details of all walled garden or email notifications; and (*iii*) abuse desk emails and walled garden contact forms that capture the communication flows between abuse department and customers. All in all, our data covers $1,688$ unique

customers who were seen to operate vulnerable devices in the ISP's consumer network between September 26th, 2017 and December 31th, 2017.

A. Vulnerability feeds

The ISP receives a daily report on vulnerable devices from the Shadowserver Foundation. These daily feeds not only identify new vulnerable devices, but also allow us to track if a device is remediated after a quarantine event or an email notification. We selected two types of vulnerabilities based on:

- *mDNS reports*: Multicast DNS (mDNS) reports provide the results from scans for publicly accessible devices that have the mDNS service accessible and answering queries. In the period of our study, a total of $1,575$ customers were found with vulnerable devices.
- *Open resolver reports:* Shadowserver open resolver reports contain information about publicly-available recursive DNS servers. Throughout our study period, we identified 113 customers with such a vulnerable device.

TABLE I: Vulnerable hosts and percentage notified

	mDNS	open resolver
# vuln. hosts	$1,575$	113
% notified	474 (30.09%)	22 (19.46%)

A daily breakdown of the number of customers reported in the feeds is shown in Figure 2. Between October 26 and November 6, 2017, the ISP did not receive any reports from Shadowserver due to server maintenance. No notifications are made during this period. Table I shows what fraction of the affected customers were notified via email or the walled garden.

The ISP does no prior filtering or inspection of the IP addresses in the Shadowserver reports before assigning treatments. This means that notifications are made irrespective of how often the IP address or customer has been seen in the reports. This is different from how most vulnerability notification experiments have been designed, where notifications are typically restricted to devices that are consistently seen over a certain period to avoid including false positive or more transient issues (e.g., [1]).

The downside of our approach, or rather the ISP's approach, is that we likely overestimate the remediation rate, as some of these devices that disappear from the reports reflect not actually remediation but transient issues. The upside is that we do not introduce selection bias. Including only those devices that are consistently seen as vulnerable over a longer period is likely to restrict the study to a non-representative subset of all vulnerable devices and their owners. Home users whose devices occasionally power off or go into standby might get excluded, for example. In other words, there is a trade-off between selection bias and overestimating the remediation rate. We decided to tolerate the latter rather than the former and to not exclude any cases that were part of Shadowserver reports and the ISP's process.

B. Notification logs

During our study period, 350 walled garden notifications were made to 327 users and 322 email-only notifications were sent to 249 users. Some users were notified more than once, sometimes with different notification types. Of all 1,688 customers in the Shadowserver reports on mDNS and open resolver, 279 also received a walled-garden notification and 3 an email-only notification for another vulnerability during the study period. For each of these notifications, we gathered (i) notification time; (ii) notification type; (iii) number of notifications made; and (iv) reason for the notification. Additionally, for walled garden notifications, we collected (i) quarantine start date; (ii) quarantine release mechanism; and (iii) quarantine removal timestamp.

C. Abuse desk logs

Notified customers can respond to the notifications via emails sent to the abuse team or, when in quarantine, via a contact form on the landing page. To better understand how users reacted to the notification and quarantine events, we gathered 564 emails from 261 users and 324 walled garden forms from 232 users.

IV. Results

In this section we evaluate the effectiveness of the notification mechanism by investigating the percentage of users that remediated in each of the following three groups: (*i*) notified and quarantined (walled garden), (*ii*) notified but not quarantined (email), and (*iii*) no action. We measure remediation via the daily reports provided by Shadowserver. There are various reasons other than remediation that might make a device disappear temporarily from the feeds, such as a temporary shutdown of the device or a disruption in the network. To conservatively estimate remediation, we check whether a vulnerable device shows up in the Shadowserver reports after the notification period, between January 1 - 31, 2018. If we do not see the device in the reports for the whole month, we assume it is remediated. This approach means we do not estimate remediation speed.

A. Measuring the impact of notifications

We first study the difference in remediation rates among the three groups: walled garden, email-only, mixed notifications and no notification (control). For this comparison, we investigate what portion of users in control group received ISP notifications for other security problems in the same observation period. This turned out to be the case for 192 users in the control group. In the same observational period, 95% of the other notifications were made for publicly-accessible devices with vulnerable NetBIOS services. Like mDNS and Open resolver, NetBIOS can be abused in amplification attacks.

While investigating the rates of remediation for the control group, we had to take into account the presence of other ISP walled garden notifications. For this reason, we divided the control group into 2 groups: (i) users who received other

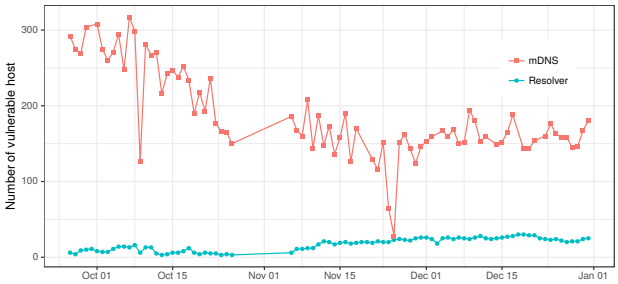


Fig. 2: Daily number of vulnerable hosts during the observation period

security notifications from the ISP; and (ii) users who received no notifications whatsoever. Table II shows the remediation rate for users in the control group who received other notifications compared to users did not receive any notifications: 96% versus 53%, respectively. A plausible explanation for this high impact of other notifications is that the typical remediation actions for NetBIOS also impact the mDNS and Open resolver vulnerabilities, e.g., disabling the DMZ or taking the device offline altogether. Table III shows that a small subset of users in the treatment groups also received other walled garden notifications. These show high remediation rates as well, but the difference is more modest compared to the other users in these groups. Note that later in this section (See section IV-E), we will present a logistic regression model that systematically controls for the impact of other notifications while estimating remediation rates for the different experimental groups.

Table II also shows that notifications for the actual vulnerability have a clear impact on its remediation. Around 87% of users in the walled-garden group remediated compared to 75% of users in the email-only group. Moreover, users that received both email and walled garden notifications on different days remediated around 81%. While the walled garden is clearly highly effective, the control group remediation rate is also surprisingly high: around 53% for the ones without any notifications and 96% for the ones that received other notifications. We will revisit this issue in the next subsection. Overall, remediation rates are high. This stands in stark contrast to most prior studies and is in the same range as the two high-profile cases of NTP amplifiers [5] and servers with the Heartbleed vulnerability [6].

B. Natural remediation

How can we make sense of the remarkably high remediation rates in the control group, even when we exclude the group who was notified for a different security issue? We consider two potential explanations: (i) transient events; and (ii) DHCP churn effects. Below, we explore the possible influence of each factor.

We first investigated role of transient events, as significant portion of the vulnerable devices reported by Shadowserver capture transient events. As discussed in III-A, we did not exclude any vulnerable devices from our study to avoid selection bias. This means that remediation rates are likely overestimated by counting transient events as remediation.

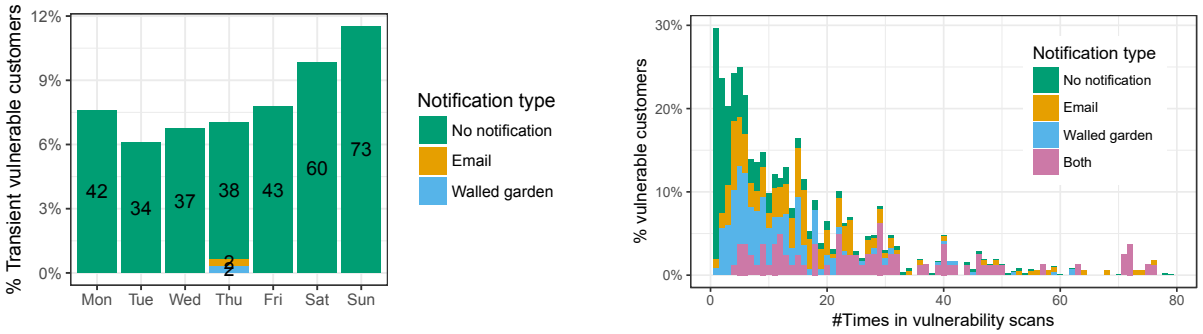
This seems to impact the control group more than other groups. As figure 3 shows, users who did not receive any notifications have a larger fraction of observations that are seen once or twice in comparison to the notified users. In total, there were 331 transient vulnerable customers of which only 4 received a notification. This is mainly due to the notification process in itself, as there is a larger fraction of transient events during the non-notification days. This is specially prominent during the weekend when the proportion of transient events increases from 20 to 40% compared to working days. This might be due to typical use cases for mDNS, namely music sharing and video streaming between devices on a home network during the weekend. As devices move from their local home networks to other networks, such as a friend's house, their mDNS functionality temporarily appears in other networks [11]. In this short period, they then appear in the vulnerability scan data. Figure 3(a) shows this pattern by visualizing the percentage of transient events, calculated as the ratio of vulnerable customers that are only reported once divided by the total amount of reported vulnerable customers per weekday.

While almost 30% of the users that did not receive notifications were seen once, only less than one percent of the notified group was seen only once. This shows that it is more likely to overestimate remediation rates of non-notified users than the ones that receive notifications. If all devices that are seen once are transient vulnerabilities, then this would already explain around half of the remediation rate of the control group.

Figure 3(b) evidences a strong correlation between the endogenous explanatory variables (i.e., the notification type) and the frequency at which a vulnerable resource is reported. Hence this frequency can be used as an instrumental variable for the consistent estimation of the remediation rates. Note that this does not invalidate the inherent randomized assignment of the notification process as this frequency is not a characteristic of the vulnerable customers, and there is no reason to believe that the characteristics of the population that present a transient vulnerability are different from those with long-lived vulnerabilities. We leverage the amount of times a vulnerable customer appears in the reports as an instrument to account for its impact on the treatment (notifications) which in turn

TABLE II: Summary statistics on the percentage of remediation according to the treatment groups and control group

	Only walled garden notification		Only email notification		Mixed notifications		Control – other walled garden notifications		Control – no notifications	
	# users	remediation rate	# users	remediation rate	# users	remediation rate	# users	remediation rate	# users	remediation rate
mDNS	225	194 (86.2%)	169	127 (75.1%)	80	65 (81.25%)	181	175 (96.6%)	920	484 (52.6)
Open resolver	22	20 (90.9%)	-	-	-	-	11	10 (90.9%)	80	48 (60.0%)
Total	247	214 (86.6%)	169	127 (75.1%)	80	65 (81.25%)	192	185 (96.3%)	1000	532 (53.2%)



(a) Percentage of transient vs. non-transient vulnerable customers per weekday

(b) Distribution of vulnerable customers appearance in the feeds

Fig. 3: Distribution of transient events per detection day and notification type

TABLE III: Remediation rates for users in different groups who also received other notifications

	Other notifications	
	#	remediation rate
Only walled garden	26	24 (92.3%)
Only email	11	10 (90.9%)
Mixed notifications	2	2 (100%)
Control	192	185 (96.3%)

impacts the remediation occurrence.

Lastly, we looked at the impact of DHCP churn on remediation rates. The ISP assigns dynamic IP address with very long DHCP lease times, typically a year. This means that a certain portion of customers has been assigned a new lease, over the course of the measurement period. This impacts our measurements for the control group differently that those for the treatment groups. The ISP's abuse department stores the IP addresses of the notified customers and we can track whether a customer has been assigned different addresses over the period of the study. The ISP does not store the IP addresses for the customers who were not notified. We had to look these up ourselves at the end of the measurement period. In other words, we could not control for churn in the control group that did not receive any security notification. As a result of this, remediation rates for 192 subscribers in control group that received security notifications for different issues were not influenced by DHCP churn. On the other hand, remediation rates for the rest of the control group might have been influenced by DHCP churn. When a vulnerable device changes its IP address, we will see the device at the old address as remediated. This means that in the control group, we will overestimate the remediation rate because of DHCP churn.

All in all, while we have no definitive explanation, these two factors help understand why the remediation rate of the control group might have been so high.Transient observations affect the remediation rates in all groups, but the control group most of all. The rate might have been further impacted by DHCP churn. This means that in reality, the difference in remediation rates between control and treatments is likely to be larger than we reported. Our conclusion that the treatments have a significantly higher impact compared to the control group is, therefore, not affected by these issues.

C. Release mechanism

We have seen that remediation rate for the walled garden group was higher than for the email-only group. The core difference between those two treatments is the incentive they provide. An email can be easily ignored, while the walled garden more forcefully compels users to act. That being said, customers can self-release from quarantine with the push of a button (see section II). In other words, if they are unable or unwilling to remediate, the walled garden does not stop them from leaving – at least not for the first two quarantine events within a month.

To see if self-release is associated with lower remediation rates, we take a closer look at the results for the different release mechanisms for users that were quarantined once: self-release, ISP-assisted release or release because the maximum quarantine period of 30 days expired.

From Table IV, we can observe that out of 236 total users, 156 (66%) used the self-release option and 86% of them remediated the vulnerability while they were in the quarantine.

This is only marginally lower than the 90% remediation rate for the 79 (33%) users who contacted ISP staff for assisted release. Just one user did not use either one of these options and his or her device was also remediated. All and all, it seems self-release did not negatively affect remediation success. The incentive mechanism worked well without being overly stringent and allowing users a speedy release and restoration of their Internet connection.

Our results show a higher remediation rate than observed by a prior study on quarantining ISP customers with a malware infection [8]. It found that quarantining incentivized 69% of 1,208 infected end-users to cleanup after the first event. The difference might be due to the fact that the vulnerabilities we studied are more transient in nature than malware infections. In both studies, assisted users showed slightly better remediation rates than the users who self-released from the walled garden.

TABLE IV: Release types and remediation

Status	1st Quarantine Event	
	Total # users	Remediation rate after Q
Self release	156	134 (85.8%)
Assisted	79	71 (89.8%)
Expired	1	1 (100%)
Total	236	206 (87.2%)

D. Measuring the impact of multiple notifications

We now take a closer look at the users who received more than one notification after they did not manage to remediate the vulnerability. Table V reports the remediation rates for these users. We separate users for whom the subsequent treatment were the same from those who received a mix of treatments. As table V demonstrates, the pattern is consistent with our earlier findings: the email-only treatment has a lower remediation success than the walled garden. Remarkably, the mixed treatments have an even higher remediation rate. We have no explanation for this result. One speculation is that it reflects how the user interprets the walled garden notification. If that treatment came first, then the subsequent email-only notification may serve as a warning that the connection might be disrupted again if the user does not act. If the email-only treatment comes first, then the subsequent walled garden action might be seen as an escalation process, compelling the user to act before further consequences are imposed.

TABLE V: Remediation after multiple notifications

	2 notifications		3 or more notifications	
	#	remediation rate	#	remediation rate
Only walled garden	11	8 (72.7%)	-	-
Only email	25	17 (68.0%)	8	6 (75.0%)
Mixed treatment	47	41 (87.2%)	33	24 (72.7%)

E. Modeling remediation occurrence

In this section we further investigate the direction and magnitude of different factors on remediation success. We investigate several observable characteristics of notifications. We use a multivariate logistic regression model that takes five explanatory (independent) variables as input:

- x_1: **Type of notification**: Categorical variable that represents the type of notification used. In our experiment we had 2 different types of notifications: (i) email and (ii) walled garden. This variable captures if one or both notification types are sent.
  - **Only email notification**: This represents users that receive only email notifications.
  - **Only walled garden notification**: This represents users notified through only walled garden notifications.
  - **Mixed notifications**: This represents users that received both email and walled garden notifications, but on different notification days.
- x_2: **Number of walled garden notifications**: Total number of walled garden notifications made per vulnerable user.
- x_3: **Number of other walled garden notifications**: Number of notifications made to a user to remediate other types of vulnerabilities. This variable captures if a user in one of the treatment of control groups received other notifications and, if so, how many. This variable allows us to distinguish two subgroups in the control group: 1,000 users who did not receive any notifications versus the 192 users who did receive a walled garden notification for another security issue (see Table II).
- x_4: **Number of email notifications**: Total number of email notifications made per vulnerable user.
- x_5: **Type of Vulnerability**: Categorical variable that shows the type of vulnerability.

These explanatory variables are included in a multivariate logistic regression model to estimate the probability of remediation occurrence. The binary logistic regression equation is explained as:

$$logit(\pi_b) = log\left[\frac{\pi_b}{1-\pi_b}\right], \quad (1)$$

where π_b is the probability of remediation within the range [0, 1] and is estimated as:

$$\pi_b = \frac{exp(\beta_0 + \sum_i \beta_i x_i)}{1 + exp(\beta_0 + \sum_i \beta_i x_i)}, \quad (2)$$

where x_i $(i = 1, \ldots, 5)$ refers to the explanatory variables; β_i is the partial regression coefficient; and β_0 is the intercept. $exp(\beta_i)$ is an odds ratio, which mirrors the strength of the association between the explanatory variables and the remediation probability. When $exp(\beta) > 1$, a positive association exists between the variables and the occurrence probability. When $exp(\beta) < 1$, a negative association exists. When $exp(\beta) = 1$, the variables are not correlated with the event.

Table VI presents the model results. We opt to fit different specifications of the model with a stepwise inclusion of the variables that impact remediation directly or indirectly.

We will first interpret the model following the standard procedure, namely via odds ratios. Next, we will translate the

TABLE VI: Coefficients of the logistic regression model for remediation

	Dependent variable: Remediation				
	(1)	(2)	(3)	(4)	(5)
x_1: Mixed notification		1.055^{***}	2.595^{***}	2.876^{***}	3.271^{***}
		(0.292)	(0.644)	(0.649)	(0.758)
x_1: Only email notification		0.695^{***}	0.695^{***}	0.871^{***}	1.192^{**}
		(0.188)	(0.188)	(0.190)	(0.374)
x_1: Only walled garden notifications		1.458^{***}	2.821^{***}	3.016^{***}	3.049^{***}
		(0.196)	(0.535)	(0.544)	(0.545)
x_2: # Walled garden notifications			-1.279^{**}	-1.330^{**}	-1.363^{**}
			(0.457)	(0.466)	(0.466)
x_3: # Other walled garden notifications				2.320^{***}	2.328^{***}
				(0.280)	(0.280)
x_4: # Email notifications					-0.234
					(0.249)
x_5: Type of vulnerability					0.289
					(0.222)
Intercept	0.687^{***}	0.412^{***}	0.412^{***}	0.153^{*}	0.130^{*}
	(0.052)	(0.059)	(0.059)	(0.063)	(0.066)
Observations	1,688	1,688	1,688	1,688	1,688
Log Likelihood	-1,076.046	-1,031.975	-1,028.358	-958.041	-956.752
Akaike Inf. Crit.	2,154.092	2,071.950	2,066.715	1,928.082	1,929.504
Note:					*p<0.05; **p<0.01; ***p<0.001

odds ratios into so-called Relative Risks, which probably are easier to understand for readers who are less familiar with odds ratios.

Exponentiating the model's coefficients gives us the odd ratios. Odds ratios express the likelihood of remediation in comparison to a reference group: the control group users (the model's intercept). (Or to be more precise: for models (2) to (4) the reference group (a.k.a. the base category) is the control group, as defined by the categorical variable x_1. In model (4), we introduce a variable (X3) to control for users in the control group who received other notifications. This does not change the reference group as such, though the intercept shifts down to accommodate the proportional influence on the log-odds of x_3. In model (5), we introduce an extra categorical variable, namely the type of vulnerability. This does change the reference group to control group users with the mDNS vulnerability. This implies that for models (2) to (4) the intercept (β_0) is the mean of the control group defined by x_1, while in model (5) the intercept is the mean of the group that constitutes the reference level for both categorical variables x_1 and x_5: control group users with the mDNS vulnerability.)

The model provides the direction and strength of the association for the predictor variables. Odd ratios above 1 mean that this factor increases the likelihood of remediation compared to the control group, while below 1 implies a decrease. We will interpret the findings based on model (5). It does not perform better than model (4), but it does enable us to look at two additional factors of interest to people designing remediation mechanisms, namely repeated email treatments and whether the type of vulnerability makes a difference. We should note, though, that the vulnerability types are actually technically similar and might show up for the same device. (As it turns out, neither variables have an observable impact on remediation.) Going from model (4) to (5), the coefficients are quite similar. The biggest change is for X1 (email-only). Even in this case, though, the coefficient of model (4) falls within the confidence interval for the coefficient of (5). Based on model (5), we can make the following observations:

- x_1: **Only Email notification**: The coefficient for email-only notifications is 1.19, which can be read as email notifications changing the log odds of remediation by 1.19. After exponentiating the coefficient, this gives us the odd ratio of 3.29 with a 95% confidence interval of [1.57,6.87] (the confidence interval is calculated by exponentiating the confidence interval for the model coefficient). In other words, the odds of remediation increase by 3.29 for users that received only email notifications, compared to the ones that did not receive any.
- x_1: **Only Walled garden notification**: By exponentiating the coefficient value, we obtain an odds ratio of 21.10 (confidence interval: [7.24,62.38]), which indicates an increase of 21.10 in the odds of remediation when notified via walled garden notifications than for not notifying.
- x_1: **Both notifications**: The odds ratio for remediation by users who received both types of notifications is 26.33 (confidence interval: [6.06,120.08]). In other words, using both walled garden and email notifications at least once in different notification days increases the odds of remediation by 26.33.
- x_2: **Number of walled garden notifications**: The coefficient for increasing the number of walled garden notifications for users who did not act upon the first notification is -1.36. This translates into an odds ratio of 0.25 (confidence interval: [0.10, 0.64]), which means

we expect to see a decrease in odds of remediation when number of walled garden events increased by one. This is consistent with our findings in Section IV, where we observed that remediation rates drop over subsequent quarantining events, indicating that these customers are less able or willing to remediate. Some common reasons why users might not act on the vulnerability notifications are discussed in Section V.

- x_3: **Number of other walled garden notifications**: The odds ratio for remediation for the 192 users in the control group who received notifications for other security issues is 10.25 (confidence interval: [6.15, 18.59]). This means there is a 10.25 increased in odds of remediation compared to those in the control group with no notifications whatsoever. This large positive impact is likely due to the fact that 95% of these other walled garden notifications were made for vulnerable NetBIOS services. The remediation steps are very similar to those for mDNS and OpenResolver. It might even concern the same device that has both vulnerable services running. Disabling the DMZ or removing the device from public access would solve both problems. Thus, we interpret the impact of X3 not so much in terms of a positive learning effect over different notifications, but rather as the effect of sharing the same – or closely related – root cause.
- x_4: **Number of email notifications**: This predictor was not significant. While email-only notifications have a positive influence on remediation, sending subsequent emails did not improve the likelihood of remediation.
- x_5: **Type of vulnerability**: Vulnerability type did not significantly influence the probability of remediation. This might be caused by the fact that both vulnerabilities (mDNS and OpenResolver) require similar actions to fix the problem.

A different way to represent these results, which might be more intuitive to some readers, is to convert the odds ratios into the so-call relative risks (RR). This captures the probability of remediation after the exposure to one of the factors as compared with the probability of remediation in the control group. The RR can be computed as:

$$RR_i = \frac{exp(\beta_i)}{1 - p_0 + (p_0 \times exp(\beta_i))}, \quad (3)$$

where p_0 represents the probability of remediation in the control group (i.e., 0.532; see Table II).

Figure 4 shows the relative risks computed from coefficients fitted in model (5) using Eq.3. Email notifications increase the probability of remediation by 30%, while walled-garden notification push up the remediation probability to 46% as compared to the control group. Adding an email notification on top of the walled garden notification only increases the probability of remediation by a non-significant 1%. (i.e., 47% probability of remediation increase compared to the control group). This suggests that the effectiveness of the mixed treatment is mainly due to the walled garden notification.

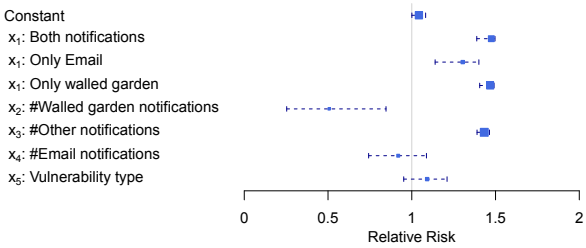


Fig. 4: Relative risks for each explanatory variable

Subsequent walled garden notifications reduce the probability of remediation by 49%. Users in the control group who received other notifications are 43% more likely to remediate than those who receive no notification whatsoever. The number of email notifications and the type of vulnerability have no significant impact on the remediation probability.

V. End User Reactions to Vulnerability Notifications

To gain insight into the user experience of a walled garden or email-only notification, we qualitatively analyzed 324 walled garden forms, as well as 564 emails to the ISP's abuse staff. This corresponds to 384 unique users, 77.4% of the 496 notified users.

We evaluated each message manually with two coders based on a subset of the themes reported by a previous study on ISP notifications [8]. New themes are added where needed. Disagreements between the two coders were adjudicated by a third coder allowing us to solve all conflicts. We found out that issues can be summarized into four categories: (*i*) expressing distrust of the notification; (*ii*) refusing to remediate; (*iii*) asking for additional help or information to solve the problem; and (*iv*) complaining about the disruption caused by the quarantining of the connection. Table VII displays the number of unique users and the percentage of all notified users with at least one message in that category or subcategory.

A. Distrusting the notification

About 1% of the notified users replied to email notifications made by the ISP to check the authenticity of the walled garden or email notifications. These users did not anticipate that their ISP would reach out to them about a vulnerable service. Interestingly, 2 users replied back to the very email they did not trust, to check the credibility of it. The other users contacted the abuse staff to check the authenticity of the quarantine landing page before they followed the suggested steps.

In another prior study, a similar degree of distrust was reported when quarantining broadband ISP subscribers with a malware infected machines [8]. This shows the importance of containing information that allows non-expert customers to reliably tell the quarantine landing page apart from a

TABLE VII: Issues raised by users in communication with the ISP

Category	# unique users
Distrusts the notification	6 (1.2%)
Unwilling or refusing to take action	91 (18.3%)
-Does not want to remove settings	80 (16.1%)
-Claims to be not vulnerable	11 (2.2%)
Requests additional information/help	215 (43.3%)
-Requests additional explanation	40 (8.1%)
-Requests additional help	169 (34.1%)
-Requests a technician	14 (2.8%)
-Request to talk with abuse desk	36 (7.3%)
-Ask for a retest	38 (7.7%)
Complains about disruptiveness	56 (11.3%)
-Cannot work due to quarantine	25 (5.0%)
-Threatens to terminate the contract	9 (1.8%)
-Cannot access devices	34 (6.9%)
Other	129 (26.0%)
No communication with abuse desk	112 (22.6%)

random phishing page. ISPs might consider personalizing the notification contents to avoid problems such as these.

B. Unwilling or refusing to take action

A noticeably high number of users did not want to act on the notifications. We distinguish two subgroups of users here. The first group does not want to change the vulnerable configuration of the device. Users argue that the suggested remediation method will prevent them from using their devices and the services that come with it, such as accessing files, playing video games with their friends, or supporting work processes. Thus, they contacted ISP to identify an alternative remediation method. In one specific case, a user mentioned that he paid a technician to set up his modem this way so that he can play games with multiple players. In another case, the user argued that disabling the DMZ and port forwarding will prevent him from monitoring his security camera from outside of his house, rendering his house less well protected. The second group contains users who complain and refuse to take action. They claim that they have been wrongly notified as they took appropriate actions before the notifications. One user claimed that they previously received another notification which also supposedly mis-identified the user's device as vulnerable.

Since a portion of subscribers were unwilling to take action, ISPs might consider withholding the quarantining for subscribers that want to keep their device configurations and suggest that they find an alternative solution to prevent abuse for amplification.

C. Requesting additional information or help

More than 40% of the users contacted the abuse desk requesting for more information or additional help to solve the problem. This category can be further divided into a few more specific themes: (i) requesting additional explanation; (ii) requesting additional help; (iii) requesting a technician; (iv) requesting to talk with abuse desk and (v) asking for a re-test. Around the first theme, users indicate that they did not properly understand the cause of the problem and requested more information from the abuse desk staff members. Several users indicated that they have been using their devices for years and wanted to know why they haven't been notified previously. Some users misunderstood the security problem and claimed to be secure with a strong login password for intruders. A few users wondered why port forwarding and enabling a DMZ are options on the ISP-issued modems, if these options are now flagged as causing security vulnerabilities. On the second theme, either users could not parse what needed to be done from the notification contents or they had questions about additional remediation methods they could try. About 3% of the users could not solve the problem by themselves and requested a paid technician from the ISP to come and fix the problem. A few indicated that they hoped the technician could find a way to fix the problem so that they can keep their configurations and devices. This rate was much lower than in two previous studies on quarantining broadband ISP subscribers with malware infections. [7], [8]. This might be because patching a vulnerable device is less complicated then cleaning up a malware-infected machine. Around 7% of the users indicated that they prefer to talk to the abuse desk employees over the phone to explain their problem. And lastly, we find out that almost 8% of the users tried to solve the problem but they were not sure about the effectiveness of the solution and they asked ISP abuse desk members to tell them whether did they managed to remediate the problem.

To reduce the number of requests made for additional help, ISPs can investigate how to improve the usability of notification content. To illustrate: a previous study on IoT malware remediation in a broadband ISP network found that providing more actionable content on the quarantine landing page reduced the percentage of requests made for additional help by half, compared to use of standard content [7].

D. Complaining about disruptiveness

During the observation period, around 11% of the notified users complained about disruptiveness of the walled garden quarantining. We further investigated the content of these messages and found several recurring themes: (i) customer states s/he cannot work due to the quarantine; (ii) customer states s/he cannot access devices; and (iii) customer threatens to terminate the contract. In the first theme, users indicated that the lack of connectivity means they cannot work from home or conduct their business properly. Around the second theme, users stated that they were out of their homes, or even out of the country, and the quarantining prevents them from accessing their network-attached storage (NAS) systems to access their backups. Finally, around 2% of users expressed anger or frustration and threatened to terminate their ISP subscription. In one case, the user additionally threatened to shame the ISP and their notification procedure on social media. A few users added they were subjected to multiple quarantine events because they could not afford to change the setting or to remove the devices that cause the vulnerability. Some users complained that quarantining users for vulnerabilities are too strong of a measure for this problem.

VI. Related Work

For many years, a large body of studies has delved into discovering vulnerabilities of different network-level entities namely websites(e.g., [12]), web applications such as CMSes [13], and web infrastructure such as servers [14]. Only in the past ten years have the security research community also put focus on studying the efficacy of notifying affected parties on remediation.

Abuse notifications: Various studies have assessed the impact of abuse notifications on cleanup of compromised websites. Notifications can be sent to the affected owners of the site or to their hosting provider. In an observational study, Li et al. used data of over 700,000 infected websites detected by Google Safe Browsing and found that direct notifications to webmasters via Google Webmaster Console increased the likelihood of cleanup by over 50% and decreased the infection lifetime by at least 62% [15]. Vasek et al. conducted an experimental study on malicious URLs submitted to the StopBadware community feeds to investigate the impact of abuse reports and how the level of detail in the reports influenced the cleanup rate [16]. They found that abuse notifications sent with detailed compromise information are cleaned up better than those not receiving a notice, 62% compared to 45% after 16 days. Notably, they found that sending a minimal report is roughly as effective as not sending at all. Cetin et al. reaffirmed that detailed notices work [17]. They concluded that while around half of all compromised websites were cleaned up after a notification to the hosting provider, sender's reputation played no statistically significant role in the clean up rates [17]. Canali at al. looked into how hosting providers handle abuse notifications [18]. They have notified 22 shared hosting providers regarding their infected webservers and observed that only 36% reacted to the abuse notifications [18]. Similarly, Nappa et al. issued abuse reports for 19 long-lived exploit servers and observed that only 7 providers took action towards cleaning up their malicious servers [19].

Vulnerability notifications: Another branch of studies have looked into how security notifications can expedite vulnerability remediation. For example, Durumeric et al. notified servers receptive to the Heartbleed vulnerability [6]. Through carrying out a controlled notification experiments two weeks after Heartbleed public disclosure, they observed that the patching rates of the notified group was 47% higher than the control group, 39.5% versus 26.8%. Kührer et al. in collaboration with CERTs, clearinghouses, and afflicted vendors notified administrators of vulnerable Network Time Protocol (NTP) servers [5]. Their results indicate 92% of NTP server were patched in 13 weeks time.

Notification mechanism: Several studies investigated specific notification mechanisms. In an earlier study, Çetin et al. investigated the usability of walled garden notifications for cleaning malware infections. The study did not include a comparison with other mechanisms or a control group, which prevented it from measuring the effectiveness of the walled garden compared to less intrusive options [8]. The observed remediation rates were around 70% after the first quarantine event, which is lower than we observed in the current study. The difference might reflect the fact that, on average, infections are harder to remediate than the studied vulnerabilities. As the prior study had no control group, we cannot see to what extent transient events might explain this difference. Such a control group was present in [7], which studied the cleanup of Mirai infections. The control group did, in fact, show a high rate of transient infection events. Overall, the study found that quarantining and notifying affected customers remediated 92% of the Mirai infections, which is in the same range as the remediation rates found in our study on vulnerabilities. Li et al. studied vulnerability notifications addressed directly to network operators and found them more effective than those send to national CERTs and US-CERT [1]. Stock et al. studied the effectiveness of large-scale email vulnerability notification campaigns. They could only reach around 6% of the affected parties. Of this small fraction, around 40% were remediated once notified [4]. Cetin et al. [2] also found email delivery rates to be poor, especially when following RFCs on how to directly contact the resource owner. Stock et al. examined the efficacy of other channels such as postal mail, social media, and phone on remediation rates. Although they resulted in marginally higher remediation rates, the gain from it do not justify the additional costs [3]. Recently, Zhang et al. looked into on the effectiveness of telephone, email, and instant message (IM) notifications within an ISP with educational institutions as main customers [20]. They conclude that IM is the most appropriate notification mode for such an ISP.

Collectively, these studies investigated the effectiveness of notifications sent to intermediaries as well as the owners of vulnerable servers and websites. However, to the best of our knowledge, there is no prior work that measured the impact of vulnerability notifications sent to end users of residential networks.

VII. Ethical Considerations

In this study, we leveraged a passively-collected dataset from an ongoing process of vulnerability and abuse handling process by the ISP. All treatments were administered by the ISP. They were existing treatments and took place within the terms of contract with their customers, so no additional consent was needed. We only added the observations from the vulnerability feeds to those treatments. The latter is not regarded as human subject research by our IRB and thus out of scope. Only the ISPs employees could see the customer information that corresponded with each observation of a vulnerable device. The study was conducted on premise at the ISP by one of the authors who was working for the ISP at the time. All raw datasets and the analysis were anonymized. Throughout the study, we followed the policies of the ISP.

VIII. Limitations

We emphasize three limitations associated with our study. First, our findings are tied to the data from a single ISP in Europe. Thus, generalizability and reproducibility of our results to other ISPs or networks are a matter for further research. Second, we only analyzed two vulnerabilities, both tied to devices being used in amplification DDoS attacks. These type of attacks usually are not directed at the vulnerable users themselves. Moreover, there is only limited media coverage of these vulnerabilities compared to, say, Heartbleed or Spectre. These factors may influence the willingness to remediate. Follow-up studies are needed to understand how this impacts remediation rates via quarantining for other vulnerabilities. Third, remediation success is measured from the scan data provided by the Shadowserver Foundation. We assume that these contain the kind of error rates normal for most large-scale scanning efforts. False negatives might lead us to incorrectly identify a host as remediated, e.g., due to temporary network disruptions. We mitigate this issue by only classifying a device as remediated if it did not appear vulnerable in Shadowserver feeds between January 1 - 31, 2018. Last, as we explained in section IV-B, there is no way to separate remediation from transient events or DHCP churn. As a result of this, we have overestimated the remediation rates, especially for the control group. This limitation should not impact our main findings – in fact, this overestimation means the difference with the treatment groups is even larger than we observed.

IX. Conclusion

We investigated the effectiveness of vulnerability notifications issued by an ISP to its customers in order to remediate devices running open DNS resolvers or mDNS services. After the three month period, we found very high remediation rates for the notified users, especially for the walled garden quarantining and notification: around 87%. These high rates also hold for users who self-released from the quarantine. The email-only notification resulted in remediation in around 75% of the cases. Few studies tracked remediation after three months and in a specific network, so it is difficult to compare these findings to prior work, but the rates are in line with those reported for the NTP amplifier campaign [5].

We explored the relatively high remediation rate for the control group: around 53%, after excluding those customers who received notifications for different vulnerabilities. Several factors cause this rate to be an overestimation. If we would remove all cases where a device was seen only once, we would end up with a remediation rate closer to what other prior studies reported [1], [6]. This would also mean that the difference in remediation rates between the notification mechanisms and the control is likely to be even larger in reality. As it stands, our analysis finds that walled garden notifications increase the probability of remediation by 46% compared to the control group. For email, we find a 30% improvement. However, sending additional walled-garden notifications to subscribers who did not act after the initial notification is associated with a decrease in the probability of remediation by 49%. This indicates certain users are unwilling or unable to remediate the vulnerability.

We have also studied the user experience of these notifications from the communications with the ISP. Quarantining vulnerable device owners is a disruptive treatment. A little over one in ten users complained about the disruption. A fraction of them even threatened to terminate the contract. It is difficult to evaluate this rate of pushback, but it seems a valid conclusion that the ISP is taking the hard road in trying to reduce the security externalities emanating from its network. Other user feedback includes a tiny fraction of users who distrusted the notifications enough to check with the ISP. Almost half of all notified users contacted the abuse department for additional information and help. Less than one in five users seemed unwilling to take action or denied having a vulnerable device to begin with. More actionable notification content might reduce the requests for help and the complaints about disruptiveness [7]. Since writing effective notification content for various vulnerabilities and infections is hard, ISPs could collaborate with researchers to conduct randomized control trials with different forms of content.

All and all, we have demonstrated that quarantining vulnerable devices is a very effective method to remediate vulnerabilities. In the setting of the ISP, email-only notifications also did much better than in Internet-wide notification experiments and control group. Reachability is likely to be much better, as is trust in the message, given that it comes from the company that users are getting service from.

The high cleanup rates achieved by quarantining and notifying vulnerable resources are comparable to, or even a bit better than, those from prior studies into walled garden notifications for compromised end user devices [7], [8]. This is remarkable, as the vulnerable devices do not pose a threat to their owners, contrary to malware-infected machines.

Notwithstanding these positive results, we do not want to overstate their contribution to solving the challenge of making large-scale vulnerability notifications more effective. The sobering observation that has to accompany our findings is that quarantining is only possible under certain conditions – e.g., the network operator needs to be contractually allowed to do so. More than contractual conditions, though, we expect that many network operators will perceive few incentives to undertake this endeavor. Walled gardens imply direct cost in terms of implementing and maintaining. Then there is the cost of time spent on notifications by the abuse handling staff. Last, but not least, there is the cost of customer pushback.

We should note that the email-only mechanism is cheaper and triggered much less customer pushback and still performed substantially better than the control group. Walled garden notifications achieved an additional 12% remediation compared to the email-only notifications. Is that additional gain worth the higher cost of the walled garden? This is a question for future work. It requires a cost-benefit analysis with the ISP, which is out of scope of the current study.

Still, we do hope that our results will encourage the community to experiment with different mechanisms in order to reach the final goal: realizing the value of large-scale vulnerability discovery for creating more secure networks.

X. Acknowledgements

This publication was supported by a grant from the Netherlands Organisation for Scientific Research (NWO), under project number 628.001.022. Also, we would like to thank Ben Stock for his generous feedback during final stages of editing this paper, the anonymous reviewers and Dennis van Beusekom for their helpful comments.

References

[1] Frank Li, Zakir Durumeric, Jakub Czyz, Mohammad Karami, Michael Bailey, Damon McCoy, Stefan Savage, and Vern Paxson. Youve got vulnerability: Exploring effective vulnerability notifications. In *USENIX Security Symposium 16*, 2016.

[2] Orcun Cetin, Carlos Ganán, Maciej Korczynski, and Michel van Eeten. Make notifications great again: learning how to notify in the age of large-scale vulnerability scanning. In *16th Workshop on the Economics of Information Security (WEIS 2017)*, 2017.

[3] Ben Stock, Giancarlo Pellegrino, Frank Li, Michael Backes, and Christian Rossow. Didn't You Hear Me?–Towards More Successful Web Vulnerability Notifications. In *The Network and Distributed System Security Symposium (NDSS)*, 2018.

[4] Ben Stock, Giancarlo Pellegrino, Christian Rossow, Martin Johns, and Michael Backes. Hey, you have a problem: On the feasibility of large-scale web vulnerability notification. In *USENIX Security Symposium 16*, 2016.

[5] Marc Kührer, Thomas Hupperich, Christian Rossow, and Thorsten Holz. Exit from Hell? Reducing the Impact of Amplification DDoS Attacks. In *USENIX Security Symposium*, 2014.

[6] Zakir Durumeric, James Kasten, David Adrian, J Alex Halderman, Michael Bailey, Frank Li, Nicolas Weaver, Johanna Amann, Jethro Beekman, Mathias Payer, et al. The matter of heartbleed. In *Proceedings of the 2014 Conference on IMC*, pages 475–488. ACM, 2014.

[7] Orçun Çetin, Lisette Altena, Carlos Gañán, Takahiro Kasama, Daisuke Inoue, Kazuki Tamiya, Ying Tie, Katsunari Yoshioka, and Michel van Eeten. Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai. In *The Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, 2019.

[8] Orçun Çetin, Lisette Altena, Carlos Gañán, and Michel van Eeten. Let me out! evaluating the effectiveness of quarantining compromised users in walled gardens. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, Baltimore, MD, 2018. USENIX Association.

[9] Messaging Anti-Abuse Working Group and others. M3AAWG best practices for the use of a walled garden. https://www.m3aawg.org/documents/en/m3aawg-best-common-practices-use-walled-garden-version-20, 2015.

[10] J Livingood, N Mody, and M OReirdan. Recommendations for the Remediation of Bots in ISP Networks (RFC 6561). *Internet Eng. Task Force*, 2012.

[11] Luiz Eduardo and Rodrigo Montoro. mdns - telling the world about you (and your device). https://www.trustwave.com/en-us/resources/blogs/spiderlabs-blog/mdns-telling-the-world-about-you-and-your-device/, 2012.

[12] Nimrod Aviram, Sebastian Schinzel, Juraj Somorovsky, Nadia Heninger, Maik Dankel, Jens Steube, Luke Valenta, David Adrian, J Alex Halderman, Viktor Dukhovni, et al. DROWN: breaking TLS using SSLv2. In *USENIX Security 16*, pages 689–706, 2016.

[13] Kyle Soska and Nicolas Christin. Automatically detecting vulnerable websites before they turn malicious. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 625–640. USENIX Association, 2014.

[14] Giancarlo Pellegrino, Onur Catakoglu, Davide Balzarotti, and Christian Rossow. Uses and abuses of server-side requests. In *International Symposium on Research in Attacks, Intrusions, and Defenses*, pages 393–414. Springer, 2016.

[15] Frank Li, Grant Ho, Eric Kuan, Yuan Niu, Lucas Ballard, Kurt Thomas, Elie Bursztein, and Vern Paxson. Remedying web hijacking: Notification effectiveness and webmaster comprehension. In *Proceedings of the 25th International Conference on World Wide Web*, pages 1009–1019. International World Wide Web Conferences Steering Committee, 2016.

[16] Marie Vasek and Tyler Moore. Do malware reports expedite cleanup? an experimental study. In *CSET*, 2012.

[17] Orcun Cetin, Mohammad Hanif Jhaveri, Carlos Gañán, Michel van Eeten, and Tyler Moore. Understanding the role of sender reputation in abuse reporting and cleanup. *Journal of Cybersecurity*, 2(1):83–98, 2016.

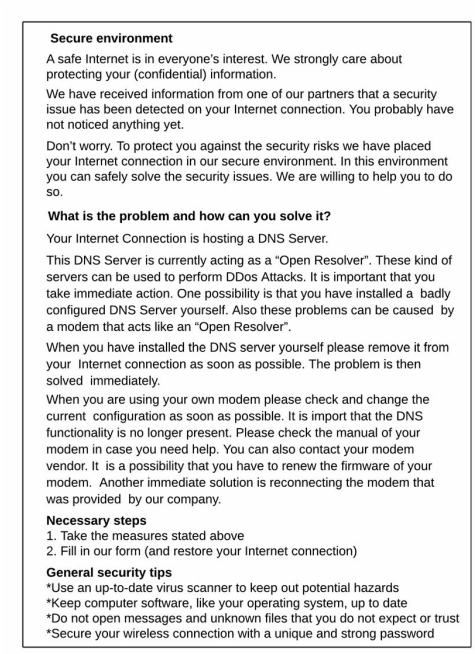
[18] Davide Canali, Davide Balzarotti, and Aurélien Francillon. The role of web hosting providers in detecting compromised websites. In *Proceedings of the 22nd international conference on World Wide Web*, pages 177–188, 2013.

[19] Antonio Nappa, M. Zubair Rafique, and Juan Caballero. Driving in the Cloud: An Analysis of Drive-by Download Operations and Abuse Reporting. In *Proceedings of the 10th Conference on Detection of Intrusions and Malware & Vulnerability Assessment*, pages 1–20, Berlin, Germany, July 2013. Springer.

[20] Jia Zhang, Haixin Duan, Wu Liu, and Xingkun Yao. How to Notify a Vulnerability to the Right Person? Case Study: In an ISP Scope. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–7. IEEE, 2017.

APPENDIX

A. Open DNS resolver walled garden notification content



Secure environment

A safe Internet is in everyone's interest. We strongly care about protecting your (confidential) information.

We have received information from one of our partners that a security issue has been detected on your Internet connection. You probably have not noticed anything yet.

Don't worry. To protect you against the security risks we have placed your Internet connection in our secure environment. In this environment you can safely solve the security issues. We are willing to help you to do so.

What is the problem and how can you solve it?

Your Internet Connection is hosting a DNS Server.

This DNS Server is currently acting as a "Open Resolver". These kind of servers can be used to perform DDos Attacks. It is important that you take immediate action. One possibility is that you have installed a badly configured DNS Server yourself. Also these problems can be caused by a modem that acts like an "Open Resolver".

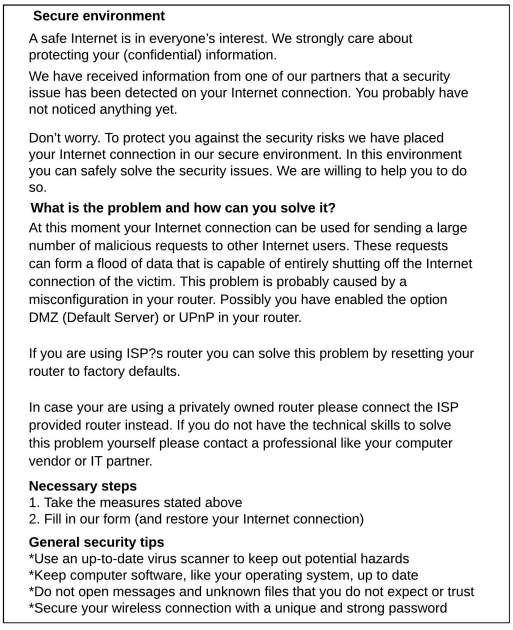
When you have installed the DNS server yourself please remove it from your Internet connection as soon as possible. The problem is then solved immediately.

When you are using your own modem please check and change the current configuration as soon as possible. It is import that the DNS functionality is no longer present. Please check the manual of your modem in case you need help. You can also contact your modem vendor. It is a possibility that you have to renew the firmware of your modem. Another immediate solution is reconnecting the modem that was provided by our company.

Necessary steps
1. Take the measures stated above
2. Fill in our form (and restore your Internet connection)

General security tips
*Use an up-to-date virus scanner to keep out potential hazards
*Keep computer software, like your operating system, up to date
*Do not open messages and unknown files that you do not expect or trust
*Secure your wireless connection with a unique and strong password

B. mDNS walled garden notification content



Secure environment

A safe Internet is in everyone's interest. We strongly care about protecting your (confidential) information.

We have received information from one of our partners that a security issue has been detected on your Internet connection. You probably have not noticed anything yet.

Don't worry. To protect you against the security risks we have placed your Internet connection in our secure environment. In this environment you can safely solve the security issues. We are willing to help you to do so.

What is the problem and how can you solve it?

At this moment your Internet connection can be used for sending a large number of malicious requests to other Internet users. These requests can form a flood of data that is capable of entirely shutting off the Internet connection of the victim. This problem is probably caused by a misconfiguration in your router. Possibly you have enabled the option DMZ (Default Server) or UPnP in your router.

If you are using ISP?s router you can solve this problem by resetting your router to factory defaults.

In case your are using a privately owned router please connect the ISP provided router instead. If you do not have the technical skills to solve this problem yourself please contact a professional like your computer vendor or IT partner.

Necessary steps
1. Take the measures stated above
2. Fill in our form (and restore your Internet connection)

General security tips
*Use an up-to-date virus scanner to keep out potential hazards
*Keep computer software, like your operating system, up to date
*Do not open messages and unknown files that you do not expect or trust
*Secure your wireless connection with a unique and strong password